

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:53:54 ; Search time 88.9 seconds
(without alignments)
1015.782 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 4202
Sequence: 1 MRNSYFKQFSAMTMAVM.....LNKKNDQDTDTVQFQIGSVF 813

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	4202	100.0	813	21	AAV44390 M. catarrhalis BAS
2	4189	99.7	813	21	AAV44391 M. catarrhalis (AT
3	1123.5	26.7	792	22	AAU03958 Neisseria meningit
4	1123.5	26.7	792	22	AAU03958 Amino acid sequenc
5	1123.5	26.7	792	22	AAU03958 Neisseria gonorrhoe
6	1115	26.5	797	22	AAU03957 Neisseria meningit
7	1115	26.5	797	22	AAU03957 Amino acid sequenc
8	1115	26.5	797	22	AAU03957 Neisseria meningit
9	1111	26.4	797	22	AAU03959 Neisseria gonorrhoe
10	1111	26.4	797	22	AAU04451 Neisseria meningit
11	1111	26.4	797	22	AAU04451 Amino acid sequenc

12	1111	26.4	797	22	AAU03958
13	1109	26.4	797	21	AAV44390
14	1104.5	26.3	792	21	AAV44391
15	1052.5	25.0	797	15	AAU03958
16	1051.5	25.0	797	15	AAU03958
17	1050.5	25.0	797	15	AAU03958
18	1047.5	24.9	793	15	AAU03958
19	1042	24.8	792	15	AAU03958
20	901	21.4	639	22	ABG17654
21	569	13.5	896	22	ABG17654
22	434	10.3	916	22	ABG17654
23	433	10.3	847	19	AAW71477
24	429	10.2	925	18	AAW57229
25	429	10.2	925	18	AAW57229
26	395	9.4	795	20	AAV17183
27	393	9.4	787	21	AAV17183
28	386.5	9.2	797	20	AAV17183
29	379.5	9.0	899	20	AAV17183
30	378.5	9.0	891	20	AAV17183
31	356.5	8.5	837	20	AAV17183
32	348.5	8.3	812	20	AAV17183
33	344.5	8.2	614	18	AAV17183
34	311	7.4	1249	22	ABG17654
35	256.5	6.1	578	21	AAV17183
36	249.5	5.9	576	21	AAV17183
37	198.5	4.7	469	21	AAV17183
38	197.5	4.7	469	21	AAV17183
39	197.5	4.7	469	21	AAV17183
40	180.5	4.3	624	21	AAV17183
41	180.5	4.3	643	21	AAV17183
42	180.5	4.3	618	21	AAV17183
43	176.5	4.2	919	21	AAV17183
44	176.5	4.2	1092	19	AAV17183
45	175.5	4.2	889	21	AAV17183

ALIGNMENTS

RESULT 1
AAV44390
ID AAV44390 standard; Protein; 813 AA.
AC AAV44390;
DT 14-MAR-2000 (first entry)
DE M. catarrhalis BAS027 polypeptide #1.
KW BAS027; OMP85; outer membrane protein; otitis media; treatment;
KW diagnosis; bacterial infection.
XX
XX Moraxella catarrhalis.
XX PN WO9963093-A2.
XX PD 09-DEC-1999.
XX PF 31-MAY-1999; 99WO-EP03822.
XX PR 03-JUN-1998; 98GB-0011945.
XX PR 08-MAR-1999; 99GB-0005304.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX Vinals-Bassols C;
XX WPI; 2000-105700/09.
XX N-PSDB; RAZ29550.
XX Novel BAS027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media


```

Db 181 parvvdiniqghfsdadliqlaiaidnknplskadrytqekivtslenlrakylmag 240
QY 241 FVRFKDAKLNEDKNRIFVEISLHEGEQYRFGOTPLGNLTQTOAELEALLKFAEE 300
Db 241 fvrfeikadkninedknrifveislhegeqyrfgqtqfignltqtaealeallkfaee 300
QY 301 GFSQAMLEQTTNNISTKFGDGGYYAQIRPVTRINDESRTVDVEYIIDPVHPVYRRINF 360
Db 301 gfsqamleqttnnistkfgdggyyaqirpvtrindesrtvdeyyidpvhpvyrrinf 360
QY 361 TGNFTQDEVLRRMQLEGALASQKQLSRARLMRGFFKHVTVDRPVPNSDDQVDV 420
Db 361 tgnftqdevlrrmqlegalasnkqqlrsarlmrfgfkhtvtdtrvpnsdpqvdv 420
QY 421 NFVVEEQSGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNAGFSRSETREVYSLGNT 480
Db 421 nfveeqsgsstiaagysqsgvtfqfdvsonnfmgtgkhvnasfrsetrevyslgnt 480
QY 481 NPFTVNGVSQSLSGYRRTKYDNKNISNYVLDYGGSLSGYGPIDENQRIISFGLNADNT 540
Db 481 npftvngvsqslsgyrrtkydnknisnyvldsygslsgypidenqrifsglnadnt 540
QY 541 KLHGGRFGISNVKQLMADGGKIQVDNNGIPDFKHDTTYNAILGWNYSLLDRPVPTQG 600
Db 541 klhggrfgisnvkqlmadggkiqvdnngipdfkhdyttynailgwnyslldrppvptqg 600
QY 601 MSHSVDLTVGFGDKTHQVVGNTYRPFIKSVKSLRGVAKLGYGNLPPYENFYAGGYGS 660
Db 601 mshsvdltvfgdkthqkvvyggnlyrpfiksvkrlgyaklgygnlppyenfyaggygs 660
QY 661 VRGYDQSSLGPRSQAYLTARRGOQTTLGEVVGNGALATFGSELILPLPKGDWIDQVRVP 720
Db 661 vrgydqsslgprsqayltarrgoqttlgevvgngalatfgselilplpkgdwidqvrpv 720
QY 721 FPIEGGVFDTCMDKQITDLTFQDPQATBQNAKAAANRPLLTQDKQLRYAGVATWY 780
Db 721 fieggvfdtcmdkqitdltfqdpqataeqnakaanrplltqdkqlryagvatwy 780
QY 781 TPIGPLSISYAKPLNKKNDQDTVOFOIGSVF 813
Db 781 tpigplsisyakplnkkndqdtvofqigsvf 813

RESULT 3
AAU03958
ID AAU03958 standard; Protein; 792 AA.
XX
AC AAU03958;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein #1.
XX
KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT FT /note= "Signal peptide"
FT Protein 22..792
FT FT /note= "Mature N. meningitidis serogroup A antigen"
XX
PN WO200138350-A2.
XX
PD 31-MAY-2001.
XX
PF 28-NOV-2000; 2000WO-IB01851.
XX
XX 29-NOV-1999; 99GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.

```

```

XX (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELSE.
XX
PI Giuliani MM, Pizsa M, Rappuoli R, Holst J;
XX
WPI; 2001-381289/40.
DR N-PSDB; AAS07278.
XX
Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
gonorrhoeae, useful in the manufacture of a medicament for treating and
preventing Neisserial bacteria infection -
XX
Claim 1; Page 66-68; 92pp; English.
XX
The sequence represents a Neisseria meningitidis serogroup A 85 kDa
antigenic protein. Neisseria meningitidis colonises the pharynx, causing
meningitis and, occasionally, septicaemia in the absence of meningitis.
This antigenic protein is useful in the manufacture of a medicament for
treating or preventing infection due to Neisseria bacteria, such as
meningitis and septicaemia. It is also useful as a diagnostic reagent for
detecting the presence of Neisseria bacteria or antibodies raised against
Neisseria, and as a reagent for raising the antibodies. The Neisserial
nucleotide sequences can be expressed in a variety of different
expression systems, for example, mammalian cells, baculoviruses, plants,
bacteria and yeast.
CC Note: There are two versions of this sequence displayed in the
CC specification (see AAU04451).
XX
SQ Sequence 792 AA;
XX
Query Match 26.7%; Score 1123.5; DB 22; Length 792;
Best Local Similarity 32.6%; Pred. No. 2.3e-67;
Matches 273; Conservative 153; Mismatches 332; Indels 79; Gaps 19;
QY 10 QVSAMTAYMYMSTHAQAADFMANDITITGLQRTVIESLQSLVPLRGLQVSVENLADG 69
Db 2 kikiasalmmgigplafadftiqdrveglrtepstvfnylpkvvgdyndthgsai 61
QY 70 VKALYATGNFSDVOYVHQEGRIYQVTRPLTAENFEGNRLIPKGLQGLKAGLAVG 129
Db 62 ikslyatgffdrvretadgqllltvierptgslntgkmlqndaklnlesfqlags 121
QY 130 QPLKQATQVMIETELTNOYISOGYVTEITVKTQMLDGNRVKLDMTFAEGKPARVVDINI 189
Db 122 qyfnqatinqavaglkceylgrgklniqtpkvtklarnrvdidiididegksakidief 181
QY 190 IGQHFSDADLIDVLAIKDNKI-NPLSKADRYTQEKLVTSLENLRKAYLNAGVFRFEIKD 248
Db 182 egqvysdrklmrqmsltgglwtlrsdrfdqrkfaadmekvtfyqnggydfriild 241
QY 249 AKLNINEDKNRIFVEISLHEGEQYRFGOTPLGNLT-YTOAELEALLKFAEGFSQAML 307
Db 242 tdiqtnedktrtkitvhhegrfrwkgvsiiegdtnevpkaeleklitmkpgkwyerqgm 301
QY 308 EOTTNNISTKFGDGGYYAQIRPVTRINDESRTVDVEYIIDPVHPVYVRRINFTGFKTQ 367
Db 302 tavlgeiqnrmgsagysvqvplpnagtktdvflhiepgarklyvneihitgnnkr 361
QY 368 DEVLRRMQLEGALASQKQLSRARLMRGFFKHVTVDRPVPNSDDQVDVNFVVEEQ 427
Db 362 devvrrlrmqmesapydtsklrskervellgfydnvqfdavplagtpdkvdlmnlalter 421
QY 428 PGSSTIAAGYSQSGGVTFQFDVSONNFMGTGKHVNAGFSRSETREVYSLGWTNPFTVN 487
Db 422 stgsidlsgavwqdtglvmsagvsqdnlfgtgksaalrasrskttingslsfcdpyftad 481
QY 488 GVSQSLSGY--YRKT---KYDNKNISNYVLDYGGSLSGYGPIDENQRIISFGLNADNTKL 542
Db 482 gvs---lgydylygkafdrprkastskqytkttagggvrmgipvteydrvfnflgaehltv 538
QY 543 HG-----GRFMGISNVKQLMADGGKITQVDNNGIPDFKHDTTYNAILGWN 587

```

Db 539 ntynkapyadfirkygtdg-----adg-----sfkg--llykgtvgwg 577
 QY 588 YSSLDPRVFTQGMHSVDLTVGF-GDK-THQKVYVQGNIVRPPFKKSVLR-----GYAK 640
 Db 578 rnktdsaswptrgyltgvnaeialpgsklqyysathnqtfflpskftlmllggevagiag 637
 QY 641 LGYG--NNLPFENFYAGGYSVGRYDOSLSGPRSQAYLTARRGQQTTLGEVW--GGNAL 696
 Db 638 -gygrtkeipfenfyggglgsrvyesgtlpgk-----vydeygekiisygnkk 686
 QY 697 ATFGSELILPLPFKGDWIDQVRPVIETGGQVFTTGMKQTDIDLTQFKDPQATAEQNAK 756
 Db 687 anvsaellfmpgkad-artvrlsfadagsv-----dgrtytaeangnknksyvenah 740
 QY 757 AANRPLLTQDKOLRYSGAGVATWTPICPLSISYAKPLNKKONDQTDVTFQIGSVF 813
 Db 741 ks-----tftnelrysagavtwlspigmkfsyayplkkpdeiqrfqfqlgttf 792
 RESULT 4
 AAB84745
 ID AAB84745 standard; Protein; 792 AA.
 AC AAB84745;
 XX
 DT 17-SEP-2001 (first entry)
 XX
 DE Amino acid sequence of a Neisseria gonorrhoeae protein.
 XX
 KW Serogroup B protein; outer membrane protein; Neisserial infection;
 XX vaccine.
 OS Neisseria gonorrhoeae.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /note= "signal peptide"
 FT Protein 22..792
 FT /note= "mature protein"
 XX
 PN WO200152885-A1.
 XX
 PD 26-JUL-2001.
 XX
 XX 17-JAN-2001; 2001WO-IB00166.
 XX
 XX 17-JAN-2000; 2000GB-0001067.
 PR 09-MAR-2000; 2000GB-0005699.
 XX
 XX (CHIR-) CHIRON SPA.
 XX
 XX Pizza M, Rappuoli R, Giuliani M;
 PI
 XX WPI; 2001-451895/48.
 DR N-PSDB; AAH42129.
 XX
 XX Composition for treating or preventing infection to, detecting, or for
 PT raising antibodies against Neisserial bacteria, comprises an N.
 PT meningitidis serogroup B outer membrane preparation and an immunogenic
 PT component -
 XX
 XX Disclosure; Page 65-67; 83pp; English.
 PS
 XX The present sequence represents a Neisseria gonorrhoeae protein. The
 CC protein is used to produce the compositions of the invention. The
 CC specification describes a composition, comprising a Neisseria
 CC meningitidis serogroup B outer membrane preparation and an immunogenic
 CC component. The immunogenic component is protein disclosed in WO99/57280,
 CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
 CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
 CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making
 CC a medicament for treating or preventing infection due to Neisserial

CC bacteria; a diagnostic reagent for detecting the presence of Neisserial
 CC bacteria or of antibodies raised against Neisserial bacteria; and/or
 CC a reagent which can raise antibodies against Neisserial bacteria. It may
 CC also be used as a vaccine.
 XX
 SQ Sequence 792 AA;
 Query Match 26.7%; Score 1123.5; DB 22; Length 792;
 Best Local Similarity 32.6%; Pred. No. 2.3e-67;
 Matches 273; Conservative 153; Mismatches 332; Indels 79; Gaps 19;
 QY 10 QVSAMTWAVMMVMTTHAQADFMANDITITGLQRTVIESLOSIVLPFRGQVSVSNQLADG 69
 Db 2 kklqiasalmmglspiafaftidqirveglrtqtepstvfnylpvkvygdtyndthgsai 61
 QY 70 VKALYATGNESDVOYVHQEGRIIYQVTERPLIAENEGNELIPKEGLEGLKNAGLAVG 129
 Db 62 klsiyatgfdvrvvetadgqlitvterptgslnitgkmlqndalkklesfglaqs 121
 QY 130 QPLKOATVQMIELETNTQYISQGYNTFIVKQTMLDGNRVKLDMTFAEGKPARVVDINI 189
 Db 122 qyfnqatlnqavaglkkeylgrgklniqitpkvklarnrvdiditidegksakitdief 181
 QY 190 IGNOHFSADLIDVLAIKDKNI-NPLSKADRYTOEKLVTSLNRAKLYLNAGFRFEIKD 248
 Db 182 egngvysdrklmrqmsltteggitwlttrsdrrfkkaqdmekvtfdygnogydfdrild 241
 QY 249 AKLMINEDKNRIFVEISLHGEQYRFGQTQFLGNT-YTQAELEALLKFAKEEFSQAML 307
 Db 242 tdiqnedktrqtkitvhieggrfrkvgsiegdnevvpkaelekillmkpgkyergqm 301
 QY 308 EQTTNNISTKFGDDGYYYAQIRPVTRINDESRTVDVEYIIDPVHPVYVRRINFTGNFKTQ 367
 Db 302 tavlgelqnrmsagayseyseisvqplpnagktvdfvlhiepgkviyneihiggnktr 361
 QY 368 DEVLREMRQLEGALASNQIKLSARLMRTGFKKHVTVTRPVNSDPDQVDVNFVVEEQ 427
 Db 362 devvrrelrgmesapydtsklqrskervellgyfdnqvfdavpagnagtdpkvdlmslter 421
 QY 428 PSGSSTIAAGYSQSGGVTFQFDVSONNFMGTGKHVNASFSRSETREYVSLGTMTPYFVFN 487
 Db 422 stgsldisagwvqdtglvmsagvsqdnlfgtgksaalrastrsktllngslftqpyftad 481
 QY 488 GVSQSLSGY--YRKT---KYDNKNISNYVLDSSYGSLSYGYPIDENQRISFGLNADNPKL 542
 Db 482 gvs--lgydiygkafdpkrkastsvkqyktttagggvrmgipvtteydrvnfglaeahitv 538
 QY 543 HG-----GREFMGISNVKQLMADGGKIQVDNNGIIPDFKHDYTYNAILGWN 587
 Db 539 ntynkapyadfirkygtdg-----adg-----sfkg--llykgtvgwg 577
 QY 588 YSSLDPRVFTQGMHSVDLTVGF-GDK-THQKVYVQGNIVRPPFKKSVLR-----GYAK 640
 Db 578 rnktdsaswptrgyltgvnaeialpgsklqyysathnqtfflpskftlmllggevagiag 637
 QY 641 LGYG--NNLPFENFYAGGYSVGRYDOSLSGPRSQAYLTARRGQQTTLGEVW--GGNAL 696
 Db 638 -gygrtkeipfenfyggglgsrvyesgtlpgk-----vydeygekiisygnkk 686
 QY 697 ATFGSELILPLPFKGDWIDQVRPVIETGGQVFTTGMKQTDIDLTQFKDPQATAEQNAK 756
 Db 687 anvsaellfmpgkad-artvrlsfadagsv-----dgrtytaeangnknksyvenah 740
 QY 757 AANRPLLTQDKOLRYSGAGVATWTPICPLSISYAKPLNKKONDQTDVTFQIGSVF 813
 Db 741 ks-----tftnelrysagavtwlspigmkfsyayplkkpdeiqrfqfqlgttf 792
 RESULT 5
 AAB23786
 ID AAB23786 standard; Protein; 792 AA.
 XX

Query Match		26.5%;	Score 1115;	DB 22;	Length 797;
Best Local Similarity		32.5%;	Pred. No. 8.8e-67;		
Matches 266;		Conservative 154;	Mismatches 361;	Indels 38;	Gaps 14;
QY	10 QVSAMTMAVMVMNSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSVENQLADG	69			
Db	2 klkqiasalmmlgisplaladftiqdirvegqrtepstvfnylpkvvgdyndthgsai	61			
QY	70 VKALYATGNFSDVOVYHQBGRYIYQVTERPLIAEINFEGRNRLPKEGLOEGLKNAGLAVG	129			
Db	62 ikslyatgfdvrvetadgqllltvierptgslnitgakmlqndaiknlesfslags	121			
QY	130 QPLKQATVQMIETELTNQVYISQGYNTEITVKOTMLDGNRVKLDMTFAEGKPARVVDINI	189			
Db	122 qyfnqatlnqavaglkceylgrgklniqitpkvtklarnrvdiditidegksakitdef	181			
QY	190 IGNQHFSDADLDVLAIKONKI-NPLSKADRYTQEKLVTSLENLRAKYLNAGFVRFEIKD	248			
Db	182 egngvysdrklmrqmsltggwtlrsnqfneqfagdmekvtfdygnngyfdfrild	241			
QY	190 IGNQHFSDADLDVLAIKONKI-NPLSKADRYTQEKLVTSLENLRAKYLNAGFVRFEIKD	248			
Db	182 egngvysdrklmrqmsltggwtlrsnqfneqfagdmekvtfdygnngyfdfrild	241			
QY	249 AKLNINEDKNRIFVEISLHGEQYRFQGTQPLGNLT-YTQAELEALLKFAEAGFSQAML	307			
Db	242 tdiqtnektktikitvhggrfwgksiegdtnevpkaeleklitmkpgkwyerqgm	301			
QY	308 EQTTNNISTKFGDGYYYAQIRPVTRINDESRVDVEYYIDPVHPVYVRRINFTGNFKTQ	367			
Db	302 tavlgeiqnrmgsagayseisvqplnaetktvdfvlhiepgkrkiyvneihitgnkntr	361			
QY	368 DEVLREMRQLEGALASNQIKLSRARLMTGFFKHVTVDRPVNSPDQVDVNFVVEEQ	427			
Db	362 devvrelrqmesapdytsklrskervellgyfdnvqfdavplagtdpkvdlmsiter	421			
QY	428 PGSSTIAAGYSQSGVTFQFDVQNNFMGTGKHVNASFSETRREVSILGNTNPFYTVN	487			
Db	422 stgsldisagwvqdtglvmsagvsqdnlfqgksaalrasrsktllngslsftdpyftad	481			
QY	488 GVSQSLSGYR--KTKYDNKNISNVYDLSYSGSLSYGYPIDENQRTSFGNLNADNTKLHGG	545			
Db	482 gvslydvvygkafprkastsikqyktttagagirmsvpvteydrvnfglvaehltvn--	539			
QY	546 RFMGISNVKQLMADGGKIQVDNNGIPDPFKHDYTTYNAILGNWYSSLDPRVFTQGNMHSV	605			
Db	540 ---tynkpkhyadffikkygktdg--tdgsfkglwlykgtvgwgrnktksalwptrgyltg	595			
QY	606 DLTVGCF-GDK-THQKVYVQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYENFYAG	656			
Db	596 naeialpgsklyqysathnqtfflpsktftlmlggevdiag-gygrtkeipffenfygg	654			
QY	657 GYGSVGVYDQSLGPRSOAYLTARGQOQTTLGEVV--GGNALATFGESEILPLPFKGDWI	714			
Db	655 glgsvrgyesgtlgbk-----vydeygekisyggnkkanvsaeellfmpgskad-a	703			
QY	715 DOVRPVIFIEGQVFTTCMDKQTDITLQFKDPQATAEQNAKANRPLLTQDKQLRYSAG	774			
Db	704 rtrvlsifadagsvwdgkytdydnssat-----ggrvqniygagnthskstftneilrysag	758			
QY	775 VGATWYTPIGPLSISYAKPLNKKQNDQTDVTFQFGSVF	813			
Db	759 gavtvlspigpmkfsayplkkkpedeqlrfqfqlgttf	797			
RESULT 8					
AAB23784					
ID	AAB23784 standard; Protein: 797 AA.				
XX					
AC	AAB23784;				
XX					
DT	12-JAN-2001 (first entry)				
XX					
DE	Neisseria meningitidis serogroup B amino acid sequence.				
XX					

KW	CG motif; complete Freund's adjuvant; phosphorothioate; immunogenic;	
KW	Neisseria antigen; Neisseria meningitidis; Neisseria gonorrhoeae;	
KW	bactericidal; antibacterial; vaccine; immunostimulatory; infection;	
KW	immune response.	
OS	Neisseria meningitidis.	
XX		
PN	WO20050075-A2.	
XX		
PD	31-AUG-2000.	
XX		
PF	09-FEB-2000; 2000WO-IB00176.	
XX		
XX	26-FEB-1999; 99US-0121792.	
PR	(CHIR-) CHIRON SPA.	
PA		
XX		
PI	Grandi G, Rappuoli R, Giuliani MM, Pizza M;	
XX		
WPI	2001-015529/02.	
XX		
PT	Immunogenic composition useful for stimulating an immune response in a	
PT	mammal against Neisseria infection, comprises Neisseria antigen and an	
PT	adjuvant composition comprising an oligonucleotide with a CG motif -	
XX		
PS	Claim 22; Page 32; 39pp; English.	
XX		
CC	The present invention describes an immunogenic composition (I)	
CC	comprising a Neisseria antigen and an adjuvant composition comprising an	
CC	oligonucleotide comprising at least 1 CG motif. Also described is an	
CC	adjuvant composition (II) comprising an oligonucleotide which comprises	
CC	at least 1 CG motif and a complete Freund's adjuvant (CFA), where the	
CC	oligonucleotide preferably comprises at least one phosphorothioate bond.	
CC	AAA92359 to AAA92385 represent specifically claimed oligonucleotides of	
CC	the present invention. (I) is useful for stimulating an immune response	
CC	in a mammal, preferably a human, against Neisseria infection, preferably	
CC	Neisseria meningitidis infection and in the manufacture of a medicament	
CC	for inducing a protective immune response in a mammal. The present	
CC	sequence represents the claimed Neisseria meningitidis serogroup B amino	
CC	acid sequence disclosed in GB-9928197.4, which is given in the present	
CC	invention.	
XX		
XX	Sequence 797 AA;	
QY		
Query Match 26.5%; Score 1115; DB 22; Length 797;		
Best Local Similarity 32.5%; Pred. No. 8.8e-67;		
Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;		
QY	10 QVSAMTMAVMVMNSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSVENQLADG	69
Db	2 klkqiasalmmlgisplaladftiqdirvegqrtepstvfnylpkvvgdyndthgsai	61
QY	70 VKALYATGNFSDVOVYHQBGRYIYQVTERPLIAEINFEGRNRLPKEGLOEGLKNAGLAVG	129
Db	62 ikslyatgfdvrvetadgqllltvierptgslnitgakmlqndaiknlesfslags	121
QY	130 QPLKQATVQMIETELTNQVYISQGYNTEITVKOTMLDGNRVKLDMTFAEGKPARVVDINI	189
Db	122 qyfnqatlnqavaglkceylgrgklniqitpkvtklarnrvdiditidegksakitdef	181
QY	190 IGNQHFSDADLDVLAIKONKI-NPLSKADRYTQEKLVTSLENLRAKYLNAGFVRFEIKD	248
Db	182 egngvysdrklmrqmsltggwtlrsnqfneqfagdmekvtfdygnngyfdfrild	241
QY	249 AKLNINEDKNRIFVEISLHGEQYRFQGTQPLGNLT-YTQAELEALLKFAEAGFSQAML	307
Db	242 tdiqtnektktikitvhggrfwgksiegdtnevpkaeleklitmkpgkwyerqgm	301
QY	308 EQTTNNISTKFGDGYYYAQIRPVTRINDESRVDVEYYIDPVHPVYVRRINFTGNFKTQ	367
Db	302 tavlgeiqnrmgsagayseisvqplnaetktvdfvlhiepgkrkiyvneihitgnkntr	361
QY	368 DEVLREMRQLEGALASNQIKLSRARLMTGFFKHVTVDRPVNSPDQVDVNFVVEEQ	427
Db	362 devvrelrqmesapdytsklrskervellgyfdnvqfdavplagtdpkvdlmsiter	421
QY	428 PGSSTIAAGYSQSGVTFQFDVQNNFMGTGKHVNASFSETRREVSILGNTNPFYTVN	487
Db	422 stgsldisagwvqdtglvmsagvsqdnlfqgksaalrasrsktllngslsftdpyftad	481
QY	488 GVSQSLSGYR--KTKYDNKNISNVYDLSYSGSLSYGYPIDENQRTSFGNLNADNTKLHGG	545
Db	482 gvslydvvygkafprkastsikqyktttagagirmsvpvteydrvnfglvaehltvn--	539
QY	546 RFMGISNVKQLMADGGKIQVDNNGIPDPFKHDYTTYNAILGNWYSSLDPRVFTQGNMHSV	605
Db	540 ---tynkpkhyadffikkygktdg--tdgsfkglwlykgtvgwgrnktksalwptrgyltg	595
QY	606 DLTVGCF-GDK-THQKVYVQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYENFYAG	656
Db	596 naeialpgsklyqysathnqtfflpsktftlmlggevdiag-gygrtkeipffenfygg	654
QY	657 GYGSVGVYDQSLGPRSOAYLTARGQOQTTLGEVV--GGNALATFGESEILPLPFKGDWI	714
Db	655 glgsvrgyesgtlgbk-----vydeygekisyggnkkanvsaeellfmpgskad-a	703
QY	715 DOVRPVIFIEGQVFTTCMDKQTDITLQFKDPQATAEQNAKANRPLLTQDKQLRYSAG	774
Db	704 rtrvlsifadagsvwdgkytdydnssat-----ggrvqniygagnthskstftneilrysag	758
QY	775 VGATWYTPIGPLSISYAKPLNKKQNDQTDVTFQFGSVF	813
Db	759 gavtvlspigpmkfsayplkkkpedeqlrfqfqlgttf	797
RESULT 8		
AAB23784		
ID	AAB23784 standard; Protein: 797 AA.	
XX		
AC	AAB23784;	
XX		
DT	12-JAN-2001 (first entry)	
XX		
DE	Neisseria meningitidis serogroup B amino acid sequence.	
XX		

KW CG motif; complete Freund's adjuvant; phosphorothioate; immunogenic; Neisseria antigen; Neisseria meningitidis; Neisseria gonorrhoeae; bactericidal; antibacterial; vaccine; immunostimulatory; infection; immune response.

XX Neisseria meningitidis.

OS

XX WO200050075-A2.

PN

XX 31-AUG-2000.

PD

XX 09-FEB-2000; 2000WO-IB00176.

PF

XX 26-FEB-1999; 99US-0121792.

PR

XX (CHIR-) CHIRON SPA.

PA

XX Grandi G, Rappuoli R, Giuliani MM, Pizza M;

XX WPI; 2001-015529/02.

DR

XX Immunogenic composition useful for stimulating an immune response in a mammal against Neisseria infection, comprises Neisseria antigen and an adjuvant composition comprising an oligonucleotide with a CG motif -

PT

XX Claim 22; Page 32; 39pp; English.

PS

XX The present invention describes an immunogenic composition (I) comprising a Neisseria antigen and an adjuvant composition comprising an oligonucleotide comprising at least 1 CG motif. Also described is an adjuvant composition (II) comprising an oligonucleotide which comprises at least 1 CG motif and a complete Freund's adjuvant (CFA), where the oligonucleotide preferably comprises at least one phosphorothioate bond. AAA92359 to AAA92385 represent specifically claimed oligonucleotides of the present invention. (I) is useful for stimulating an immune response in a mammal, preferably a human, against Neisseria infection, preferably Neisseria meningitidis infection and in the manufacture of a medicament for inducing a protective immune response in a mammal. The present sequence represents the claimed Neisseria meningitidis serogroup B amino acid sequence disclosed in GB-9928197.4, which is given in the present invention.

CC

XX Sequence 797 AA;

SQ

Query Match 26.5%; Score 1115; DB 22; Length 797;

Best Local Similarity 32.5%; Pred. No. 8.8e-67;

Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTMAVMVMNSTHAQAADFMANDDITITGLQRTVIESLQSVLPFRLGQVSVENQLADG 69

Db 2 ktkqasalmgisplaladftiqdirvegqrtepstvfnylpkvvgdyndthgsal 61

QY 70 VKALYATGNFSDVOVYHQBGRYIYQVTERPLIAEINFEGRNRLPKEGLOEGLKNAGLAVG 129

Db 62 ikslyatgfdrrvetadgqllltvierptgslnitgakmlqndaiknlesfslags 121

QY 130 QPLKQATVQMIETELTNQVYISQGYNTEITVKOTMLDGNRVKLDMTFAEGKPARVVDINI 189

Db 122 qyfnqatlnqavaglkceylgrgklniqitpkvtklarnrvdiditidegksakitdef 181

QY 190 IGNQHFSDADLDVLAIKONKI-NPLSKADRYTQEKLVTSLENLRAKYLNAGFVRFEIKD 248

Db 182 egngvysdrklmrqmsltggwtlrsnqfneqfagdmekvtfdygnngyfdfrild 241

QY 249 AKLNINEDKNRIFVEISLHGEQYREGQTQFLGNLT-YTQAELEALLKFAEGFSQAML 307

Db 242 tdiqtnektktikitvhggrfwgksiegdtnevpkaeleklitmkpgkwyerqgm 301

QY 308 EQTTNNISTKFGDGYYYAQIRPVTRINDESRVDVEYYIDPVHPVYVRRINFTGNFKTQ 367

Db 302 tavlgeiqrmgsagayseisvqplpnaetktvdfvlhiepgkrkiyvneihitgnkntr 361

QY 368 DEVLREMRQLGALASNOIKQLSRARLWRTGFFKHVTVDRPVNSPOQVDVNFVVEEQ 427

Db 362 devvrelrqmesapdytsklrskervellgyfdnvqfdavplagtdpkvdlnmsalter 421

QY 428 PGSSTIAAGYSQSGVTFQFQVQNNEMGTGKHVNASFSETRREVSILGNTNPFYTVN 487

Db 422 stgsldlsagwqdtglvmsagvsqdnlfgtkksaalrasrskttinglsftdpyftad 481

QY 488 GYSQSLSGYR--KTKYDNKNISNVYDLSYSGSLSYGYPIDENQRTSFGNLNADNTKLHGG 545

Db 482 gvslydvvygkafprkastsikqktttagagirmsvpvteydrvnfglvaehltvn-- 539

QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDPFKHDYTTNAILGNWYSSLDPRVFTQGNHSHV 605

Db 540 ---tynkpkhyadffikkygktdg--tdgsfkglwlykgtvgwgrnktksalwptrgyltg 595

QY 606 DLTVGf-GDK-THQKVYQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYENFYAG 656

Db 596 naeialpgsklyysathnqtfflpsktftlmlggevdiag-gygrtkeipffenfygg 654

QY 657 GYGSVGVYDQSLGPRSOAYLFPARRGQQTGLGEVV--GGNALATFGESEILPLPFKGDWI 714

Db 655 glgsvrgyesgtlgbk-----vydeygekisyggnkkanvsaeellfmpgskad-a 703

QY 715 DOVRPVITEGQVFTTCMDKQTDITLQFKDPQATAEQNAKANPPLTQDKQLRYSAG 774

Db 704 rtrvlsifadagsvwdgkytdydnssat-----ggrvqniygagnthkstftcneilrysag 758

QY 775 VGATWYTPIGPLSISYAKPLNKKQNDQTDVTFQFGSVF 813

Db 759 gavtvlspigpmkfsayplkkkpedeqlrfqfqlgttf 797

QY 368 DEVLRRMRQLEGALASNQIKQLSRARLMRTGFFKHVTVTRPVPNSPDQVDVNFVEEQ 427
 Db 362 devrrrelrqlmesapydtsklrskervellgyfdnvqfdaavplagtpdkvdlmslter 421
 QY 428 PSGSSTIAAGYSQGGVTFQFDVSONFMGTGKHVNASFSSRSETRVYSLGTMNPYFTVN 487
 Db 422 stgslldisagwvqdtglvmsagvsqdnlfgtgksaalrskrktlsgslsftdpyftad 481
 QY 488 GVSQSLSGYYR--KTKYDNKNISNVYLDSSYGGSLSYGYPIDENORISFGLNADNTKLHGG 545
 Db 482 gvslygydvgykafprkastsikyktttagagirmsvpvteydrvnfvlvaehltvn-- 539
 QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDPKHDYTYNAILGNWYSSLDPRVPTQGMHSV 605
 Db 540 ---tynkaphyadikkygktgdg--tdgsfkgwlygtvgwgrnktksalwptrgyitgv 595
 QY 606 DLTVGF-GDK--THQKVYVQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYENFYAG 656
 Db 596 naeialpgsklqyysathnqtwtfflpskftlmlggevgiag-gygrtkeipffenfygg 654
 QY 657 GYGSVRGYDQSSSLGRSQAYLTARRGQOITLGEVY--GGNALATFGSELILPLPKGDWI 714
 Db 655 glgsvrgyesgtlqpk-----vydeygekiysygnkkanvsaeallfmpgagkd-a 703
 QY 715 DOVRPVFIEGGQVDFDTGMDKQITDLTQFKDPOATAEQNAKAAANRPLLTDQKOLRYSG 774
 Db 704 rtrvlsifadagsvwdgktydndssat-----ggrvgniyyagdnthkstkftneirysag 758
 QY 775 VGATWYPIGPGLSISYAKPLNKKQNDQTDVQFOIGSVF 813
 Db 759 gavtwslpgmksyapyklkpedelqrfqfqlgttf 797

RESULT 9
 AAU03959
 ID AAU03959 standard; Protein; 797 AA.
 XX AAU03959;
 XX
 DT 23-OCT-2001 (first entry)
 XX
 DE Neisseria gonorrhoeae antigenic protein.
 XX
 KW Antigenic protein; pharynx; meningitis; septicaemia; mammalian cell;
 KW bacterial infection; baculovirus; yeast; Neisseria meningitidis.
 OS Neisseria gonorrhoeae.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT Protein /note= "Signal peptide"
 FT /note= "Mature N. gonorrhoeae antigen"
 XX
 PN W0200138350-A2.
 XX
 PD 31-MAY-2001.
 XX
 XX 28-NOV-2000; 2000WO-IB01851.
 XX
 PR 29-NOV-1999; 99GB-0028197.
 PR 09-MAR-2000; 2000GB-0005698.
 XX
 XX (CHIR-) CHIRON SPA.
 PA (STAT-) STATENS INST FOLKEHELSE.
 PA
 XX Giuliani MM, Piza M, Rappuoli R, Holst J;
 XX
 XX WPI; 2001-381289/40.
 DR N-PSDB; AAS07279.
 XX
 PT Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
 PT gonorrhoeae, useful in the manufacture of a medicament for treating and

preventing Neisserial bacteria infection -
 Claim 1; Page 37-39; 92pp; English.
 XX
 CC The sequence represents a Neisseria gonorrhoeae 85 kDa antigenic protein.
 CC N. gonorrhoeae is closely related to N. meningitidis, which colonises the
 CC pharynx, causing meningitis and, occasionally, septicaemia in the absence
 CC of meningitis. This antigenic protein is useful in the manufacture of a
 CC medicament for treating or preventing infection due to Neisseria
 CC bacteria, such as meningitis and septicaemia. It is also useful as a
 CC diagnostic reagent for detecting the presence of Neisseria bacteria or
 CC antibodies raised against Neisseria, and as a reagent for raising the
 CC variety of different expression systems, for example, mammalian cells,
 CC baculoviruses, plants, bacteria and yeast.
 XX
 SQ Sequence 797 AA;
 Query Match 26.4%; Score 1111; DB 22; Length 797;
 Best Local Similarity 32.4%; Pred. No. 1.6e-66;
 Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;
 QY 10 QVSAMTNAVMMVMTTHAQAAADPMANDIITGLQRTVIESLSQSLVPLFRGQVYSENLADG 69
 Db 2 kklqasalmvlgisplaladftlqdirvegqlrtepstvfnylpvkvgdyndthgsai 61
 QY 70 VKALYATGNFSDVOVYHOEGRIIYQVTERPLIAETNFEGNRLIPKEQLEGKLNAGLAVG 129
 Db 62 kkslyatgffdvrvetadgqllltvierptgslnigakmlqndaikklesfglaqs 121
 QY 130 QPLQOATVQMIETELTNOYISQYYNTEITVQTMLDGNRVKLDMTFAEGKPARVVDINI 189
 Db 122 qyfqatlnqavaglkceylgrgklniqtpkvtklarnrvdiditidegksakitdef 181
 QY 190 IGNOHFSADLIDVLAIKDKNI-NPLSKADRYTQKLVTSLENLRAKYLNAGFVFEIKD 248
 Db 182 egngvysdrklmrqmsitdeggiwtlrsnqfneqfagdmekvtdfygnngyfdfrild 241
 QY 249 AKLNEDKNRIFVEISLHEGEYRFGOTQFLGNTL-VTQAELEALLFKAEESQAML 307
 Db 242 tdiqtnedkktqtkitvhggrfrwkvsgiednevpkaeleklmlmkpgkyvyrqgm 301
 QY 308 EQTTNNISTKFGDGYIAQIRPVTRINDESTRVDVEYIIDPVHPVYVRRINFTGNFKTQ 367
 Db 302 tavigeiqnrmgsagayseisvqplpnaetktdfvlhiepggrkiyvnehitgunktr 361
 QY 368 DEVLRRMRQLEGALASNQIKQLSRARLMRTGFFKHVTVTRPVPNSPDQVDVNFVEEQ 427
 Db 362 devrrrelrqlmesapydtsklrskervellgyfdnvqfdaavplagtpdkvdlmslter 421
 QY 428 PSGSSTIAAGYSQGGVTFQFDVSONFMGTGKHVNASFSSRSETRVYSLGTMNPYFTVN 487
 Db 422 stgslldisagwvqdtglvmsagvsqdnlfgtgksaalrskrktlsgslsftdpyftad 481
 QY 488 GVSQSLSGYYR--KTKYDNKNISNVYLDSSYGGSLSYGYPIDENORISFGLNADNTKLHGG 545
 Db 482 gvslygydvgykafprkastsikyktttagagirmsvpvteydrvnfvlvaehltvn-- 539
 QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDPKHDYTYNAILGNWYSSLDPRVPTQGMHSV 605
 Db 540 ---tynkaphyadikkygktgdg--tdgsfkgwlygtvgwgrnktksalwptrgyitgv 595
 QY 606 DLTVGF-GDK--THQKVYVQGNIRYRPFIKKSVLR-----GYAKLGYG--NNLPFYENFYAG 656
 Db 596 naeialpgsklqyysathnqtwtfflpskftlmlggevgiag-gygrtkeipffenfygg 654
 QY 657 GYGSVRGYDQSSSLGRSQAYLTARRGQOITLGEVY--GGNALATFGSELILPLPKGDWI 714
 Db 655 glgsvrgyesgtlqpk-----vydeygekiysygnkkanvsaeallfmpgagkd-a 703
 QY 715 DOVRPVFIEGGQVDFDTGMDKQITDLTQFKDPOATAEQNAKAAANRPLLTDQKOLRYSG 774

Db 704 rtvrlslfadagvwdgktyddnssat-----ggrvqnlygagnthkstftneirysag 758

QY 775 VGATWTPIGPLSISYAKPLNKKQNDQTDVQFOIGSVF 813

Db 759 gavtwsplgpmkfsayayplkkpdeiqrfqfqtgtf 797

RESULT 10

AAU04451

XX AAU04451 standard; Protein; 797 AA.

AC AAU04451;

XX 23-OCT-2001 (first entry)

XX Neisseria meningitidis serogroup A antigenic protein #2.

XX Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;

KW bacterial infection; baculovirus; yeast.

XX Neisseria meningitidis.

OS Key Location/Qualifiers

PH Peptide 1..21

FT /note= "Signal peptide"

FT Protein 22..797

FT /note= "Mature N. meningitidis serogroup A antigen"

PN W0200138350-A2.

XX 31-MAY-2001.

XX 28-NOV-2000; 2000WO-IB01851.

XX 29-NOV-1999; 99GB-0028197.

PR 09-MAR-2000; 2000GB-0005698.

XX (CHIR-) CHIRON SPA.

PA (STAT-) STATENS INST FOLKEHELSE.

PI Giuliani MM, Pizza M, Rappuoli R, Holst J;

XX WPI; 2001-381289/40.

XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria gonorrhoeae, useful in the manufacture of a medicament for treating and preventing Neisserial bacteria infection -

PT Claim 1; Page 39-40; 92pp; English.

XX The sequence represents a Neisseria meningitidis serogroup A 85 kDa antigenic protein. Neisseria meningitidis colonises the pharynx, causing meningitis and, occasionally, septicaemia in the absence of meningitis. This antigenic protein is useful in the manufacture of a medicament for treating or preventing infection due to Neisseria bacteria, such as meningitis and septicaemia. It is also useful as a diagnostic reagent for detecting the presence of Neisseria bacteria or antibodies raised against Neisseria, and as a reagent for raising the antibodies. The Neisserial nucleotide sequences can be expressed in a variety of different expression systems, for example, mammalian cells, baculoviruses, plants, bacteria and yeast.

CC Note: There are two versions of this sequence displayed in the specification (see AAU03958).

XX Sequence 797 AA;

XX

Query Match 26.4%; Score 1111; DB 22; Length 797;

Best Local Similarity 32.4%; Pred. No. 1.6e-66;

Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTAVMMVMSHAQAADFMANDITITGLQRVTIESLOSVLFPRLGQVVSQNLADG 69

XX

Db 2 klkqiasalmvlgisplaladftiqdirvegqrtepstvfnylvpkvgtndthgsai 61

QY 70 VKALYATGNFSDVOYVYHQESRIYQVTERPLAEINFEGRNLIPKLEGLKAGLAVG 129

Db 62 ikslyatgffddrvvetadgqilltvtierptgslnftgakmlqndafknlesfags 121

QY 130 OPLKQATVQMIETELTNOYISQYYNTEITVKTMLDGNRVKLDMTFAGCKPARVVDINI 189

Db 122 qyfnqetlnqavaglkceylgrgklniqtpkvtklarnrvdidditidegksakidief 181

QY 190 IGNOHFSADLDLVLAKNKI-NPLSKADRYTOEKLVTSLENLAKYLNAGVREIKD 248

Db 182 egqvysdrklmrqmsltteggitwlttranfneqfagdmekvtfygnngydfriid 241

QY 249 AKLINEDKNRIFVEISLHEGQYRFGQFQFLGNLT-YTQAELEALLKFAEGFQOAML 307

Db 242 tdiqtnedktktiktvtvheggrfwgkvsiegdtnevpkaelekiltmkpgkwyerqgm 301

QY 308 EOTTNNISTKFGDDGYVYQAIRPVTIRINDESRTVDVEYIDPVHPVYVRRINFTGNFKTQ 367

Db 302 tavlgeiqnrmgsagayayseisvqplpnaetktvdfvlhiepgklyvneihtgnktr 361

QY 368 DEVLREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRVPVNSPOQDVNFVEEQ 427

Db 362 devvrrelrqmesapydtsklgrskervellgyfdnvqfdavplagtpdkvdlmnlter 421

QY 428 PGGSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNASFSSRSTREVSILGTMTPYFTVN 487

Db 422 stgsidlsagvwdtgi vmsagvsqdnlfgtksaalrasrsktlingslsfcdpyftad 481

QY 488 GVSQSLSGYR--KTKYDNKNISNVYLDYSGGSLSYGYPIDENQRIISFGLNADNTKLHGG 545

Db 482 gvslydyvgkafprkastsikyktttagagirmsvpyteydrvnfvlvaehtv-- 539

QY 546 RFWGISNVKQLMADGKIQVDNNGIPDFKHDTTYTNAILGNVSSLDTRVPVFPQGMHSV 605

Db 540 ---tynkepkyadfkikygktidg--tdgsfkglwlykgvgnrktidsalwptrgyltg 595

QY 606 DLTVGf-GDK--THQKVYVYOGNIYRPEIKKSLVR----GYAKLGYG--NNLPYENFYAG 656

Db 596 naeialpgsklqyysathnqtqtfplsktftlmkggevglag-gygrtkeipfenfyg 654

QY 657 GYGSVRGYDQSSLPSPRSQAYLTARRQQOITLGEV--GGNALATFGSELIPLPFGDWI 714

Db 655 glgsrvygesglgpk-----vydeygekisygnkkanvsaeilfmpgagd-a 703

QY 715 DQVRPVIFIEGQGVDTTGMKQTDLTQFKDPQATAEQNAKAAANRPLLTQDKQLYSAG 774

Db 704 rtvrlslfadagvwdgktyddnssat-----ggrvqnlygagnthkstftneirysag 758

QY 775 VGATWTPIGPLSISYAKPLNKKQNDQTDVQFOIGSVF 813

Db 759 gavtwsplgpmkfsayayplkkpdeiqrfqfqtgtf 797

RESULT 11

AA884746

XX AA884746 standard; Protein; 797 AA.

AC AA884746;

XX 17-SEP-2001 (first entry)

XX Amino acid sequence of a Neisseria serogroup A protein.

XX Serogroup A protein; outer membrane protein; Neisserial infection;

KW vaccine.

XX Neisseria meningitidis.

OS Neisseria meningitidis.

PH Key Location/Qualifiers

FT Peptide 1..21

FT /note= "signal peptide"

QY 249 AKLNINEDKNRIFVEISLHEGEQYRFGQTQFLGNLT-YTOAELEALLKFAEEGFSQAML 307
 Db 242 tdiqtnedkqtikitvheggrfwkvslegdtnepvkaeleklitmkpgkwyerqgm 301
 QY 308 EOTTNNISTKFGDDGYYIAQIRPVTRINDESRVTVDVEYIDPVHPVYVRRINFTGNKFTQ 367
 Db 302 tavgeiqnrmgsagayseisvqplnaeetktdvflhiepggrkiyvneihitgnnkt 361
 QY 368 DEVLRREROLEGALASNQIKQLSRARLMRTGFFKHVHTVDRPVPNSPDQVDVNFVVEEQ 427
 Db 362 devvirelrmesapdytskqrskervellgyfdnqfdaavpdkvldlmsiter 421
 QY 428 PSGSSTIAAGYSQSGGVTFQDVSONNPMFGTKHVNASFSSRSETRVYSLGTMNPFTVN 487
 Db 422 stgslidisagvwdtglmvsagvsqdnlfgtksaalrasrsktllngslsftdpyftad 481
 QY 488 GVSQSLSGYR--NKTQDNKNISNVLDSSYGSLSYGYPIDENORISFGLNADNTKLHG 545
 Db 482 gvslygydvgykafdrpkastiskyktttagdrmsvptveydrvnfvglaehltn-- 539
 QY 546 RFMGISNVKQLMADGGKIQVDNNIPDFKHDYTYNAILGWNYSLSLDRPVPPTQGMHSV 605
 Db 540 ---tynkapkhyadfikkygktgdg--tdgskfwlykgtvgwgrnktksalwptrgyitgv 595
 QY 606 DLTVGF-GDK--THOKVYVQGNIPYRPFIKKSVLR-----GYAKLGYG--NNLPFFYENYAG 656
 Db 596 naelalp9sklqyysathnqtwfplsktftlmiggevgiag-gygrtkelpffenyf 654
 QY 657 GYGSVRGYDQSSLGPRSOAYLTARRGQOTTLGEVV--GGNALATFGSELILPLPFKGDWI 714
 Db 655 gl9svryesagtlgpk-----vydeygekisvggnkanvsaellfmpgpkad-a 703
 QY 715 DQVRPVIFIEGGQVDFDTGMDKQIDTLQFKDPOATAEQNAKANRPLLTQDKOLRYSAG 774
 Db 704 rrtvlsifadagsvwdgktydnnssat-----ggrvqniygagethksthftnelrysag 758
 QY 775 VGATWYPIGLPISYAKPLNKKQNDQTDVQFQIGSVF 813
 Db 759 gavtwisplgpmkfrayaypkkkpedelqrfqqlgttf 797

RESULT 14
 AAY84946
 ID AAY84946 standard; Protein; 792 AA.
 XX AAY84946;
 AC AAY84946;
 XX 21-AUG-2000 (first entry)
 XX Amino acid sequence of outer membrane protein (omp) 85.
 DE Outer membrane protein; omp85; gonococcal infection; symptomatic disease;
 KW meningococcal infection; protective immune response; vaccine.
 XX Neisseria gonorrhoeae.
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /note= "signal peptide"
 FT 1..21
 XX WO200023595-A1.
 XX 27-APR-2000.
 XX 22-OCT-1998; 98WO-US22352.
 XX 22-OCT-1998; 98WO-US22352.
 XX (UYMO-) UNIV MONTANA.
 PA Judd RC, Manning SD;
 XX Judd RC, Manning SD;
 XX

DR WPI; 2000-339694/29.
 XX N-PSDB; AAA15155.

PT New isolated outer membrane protein 85 of Neisseria gonorrhoeae and N.
 PT meningitidis useful for vaccine, therapeutic and diagnostic
 PT compositions for gonococcal or meningococcal infections -
 XX Claim 1; Fig 2; 98pp; English.

XX The present sequence represents an outer membrane protein (omp) 85 of
 CC Neisseria gonorrhoeae. The omp polypeptides and polynucleotides are
 CC useful in compositions for use in the prevention, treatment and diagnosis
 CC of non-symptomatic gonococcal infection or meningococcal infection and
 CC symptomatic disease. They are also useful for the detection of
 CC hybridisation complexes. Antigens and antibodies specific omp proteins
 CC also provide diagnostic, therapeutic and prophylactic compositions for
 CC the treatment or prevention of the infections described above. The
 CC antibodies are useful for inducing a protective immune response in
 CC humans or animals with N. gonorrhoeae, N. meningitidis, or other
 CC Neisseria species. The proteins, antibodies and polynucleotide
 CC sequences of the present invention may also be used in the screening
 CC and development of chemical compounds such as drugs or vaccines.

XX Sequence 792 AA;

Query Match 26.3%; Score 1104.5; DB 21; Length 792;
 Best Local Similarity 32.6%; Pred. No. 4.4e-66;
 Matches 268; Conservative 155; Mismatches 350; Indels 49; Gaps 17;

QY 10 QVSAMTAVMMVMSTHAQADFMANDITITGLQRTVIESLQSVLPFRQGVVSNQLADG 69
 Db 2 kklqiasalmmglisplafadftiqdirveglqrtepstvfnylpvkvgdyndthgsai 61
 QY 70 VKALYATGNSDVQVYHQBEGRIIYQVTERPLIAEINFEGRNLIPKEGLQEGKNAVLAVG 129
 Db 62 ikslyatgfdvdrvetadgllltivcptglsintgamlqndalknlesfglas 121
 QY 130 QPLKQATVQIETELTNQYISQGYINTEITVKQTMLDGNRVKLDWMTFAEGPARVVDINI 189
 Db 122 qyfnqatinqavaglkkeeylgrgklniqtpkvtklarnrvdiditidegksakitdief 181
 QY 190 IGNOHFSDADLIDVLAIKDKNI-NPLSKADRYTOEKLVTSLENRKAYLNAGVRFETKD 248
 Db 182 egngqvysdrklmrqmsltedgiwtlrsdrfdqrkfaqdmekvtdfyngnygdfdrill 241
 QY 249 AKLNINEDKNRIFVEISLHEGEQYRFGQTQFLGNLT-YTOAELEALLKFAEEGFSQAML 307
 Db 242 tdiqtnedkqtikitvheggrfwkvslegdtnepvkaeleklitmkpgkwyerqgm 301
 QY 308 EOTTNNISTKFGDDGYYIAQIRPVTRINDESRVTVDVEYIDPVHPVYVRRINFTGNKFTQ 367
 Db 302 tavgeiqnrmgsagayseisvqplnaeetktdvflhiepggrkiyvneihitgnnkt 361
 QY 368 DEVLRREROLEGALASNQIKQLSRARLMRTGFFKHVHTVDRPVPNSPDQVDVNFVVEEQ 427
 Db 362 devvirelrmesapdytskqrskervellgyfdnqfdaavpdkvldlmsiter 421
 QY 428 PSGSSTIAAGYSQSGGVTFQDVSONNPMFGTKHVNASFSSRSETRVYSLGTMNPFTVN 487
 Db 422 stgslidisagvwdtglmvsagvsqdnlfgtksaalrasrsktllngslsftdpyftad 481
 QY 488 GVSQSLSGYR--NKTQDNKNISNVLDSSYGSLSYGYPIDENORISFGLNADNTKL 542
 Db 482 gvs---lgydiygkafdrpkastiskyktttagggrvmgipvtveydrvnfvglaehltn 538
 QY 543 HGRFEMGISNVKQLMADGGKIQVDNNIPDFKHDYTYNAILGWNYSLSLDRPVPPTQOMS 602
 Db 539 n-----tynkapkryadfikkygktgdg--adgsfkglykgtvgwgrnktksalwptrgyl 592
 QY 603 HSYDLTVGF-GDK--THOKVYVQGNIPYRPFIKKSVLR-----GYAKLGYG--NNLPFFYENF 653
 Db 593 tgvnaeiaipgsklqyysathnqtwfplsktftlmiggevgiag-gygrtkelpffenyf 651

Qy	654	YAGYGSVGRGYDSSLGPRSOAYLTARRGOQTTLGEVV--GGNALATFGSEILPLPFPKG	711
Db	652	YGGIIGSVRGVYSGTLLGPK-----vyydeygekisyggnkkanvsaellfmpgag	701
Qy	712	DWIDQVRPVIFIEGGQVEDTGMKDQKQIDLTQFKDQQAETQNAKAANRPLLTQDKQLRY	771
Db	702	d-actvrlsfadagsvw-----dgrtytaaengnnkksvysenahks-----tftnelry	750
Qy	772	SAGVGATWYPIGPISLISYAKPLNKKQNDQTDVQFQIGSVF	813
Db	751	saggavtwlspgmkfviyplkkpdeiqrfqqlgttf	792
RESULT	15		
AA	3755		
ID	AA	53755 standard; Protein; 797 AA.	
XX			
AC	AA	53755;	
XX			
DT	08-DEC-1994	(first entry)	
XX			
DE		H. influenzae b Eagen D15 sequence.	
XX			
KW		Vaccine; passive immunization; vector; antiserum; diagnosis; D15;	
KW		OMP; outer membrane protein; Hib.	
XX			
OS		Haemophilus influenzae type b Eagen strain.	
XX			
PN		W09412641-A.	
XX			
PD	09-JUN-1994.		
XX			
PF	23-NOV-1993;	93WO-CA00501.	
XX			
PR	23-NOV-1992;	92GB-0024584.	
XX			
PA	(CONN-) CONNAUGHT LAB LTD.		
XX			
PI	Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;		
PI	Yang Y;		
XX			
DR	WPI; 1994-200269/24.		
XX			
DR	N-PSDB; AAQ66199.		
XX			
PT		Nucleic acid encoding D15 outer membrane protein - esp. of	
PT		Haemophilus influenzae, and related proteins, vectors, antisera	
PT		etc. useful in vaccines, for diagnosis and for passive	
PT		immunisation.	
XX			
PS	Disclosure; Fig. 1B; 161pp; English.		
XX			
CC		Outer membrane protein (OMP) D15 genes were isolated by screening	
CC		chromosomal libraries of H. influenzae type b (Hib) strains Ca,	
CC		Eagan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK	
CC		12085. Nucleotide sequences were determined for the D15 genes	
CC		(AAQ66198-202) and the corresponding aa sequences were derived	
CC		(AAQ53754-58). D15 OMP can be produced easily and on a large scale,	
CC		free of other antigens and lipooligosaccharides, by recombinant DNA	
CC		methods using the isolated genes.	
XX			
SQ	Sequence	797 AA;	

Query Match	25.08;	Score 1052.5;	DB 15;	Length 797;
Best Local Similarity	31.68;	Pred. No. 1.4e-62;		
Matches 260;	Conservative 160;	Mismatches 333;	Indels 69;	Gaps 20;
QY	23	STHAQADEFWANDITITGLQRTVTTESLQSVLPFRLGQVVSENQLADGVKALYATGTFNSDV	82	
Db	14	tttfagpfvakdlrvdvgdgldqqiravpavqgrvtndvanlrvslfvsgrfdv	73	
QY	83	QVHQEGRI-IYQVTERPLTAIEINFEGNRLRLLPKKEGLKNAGLAVGQPLKQATVOMIE	141	

Db	74	ka-hqegdvlvsvvaksilsdsvikngsviptealkqnl.dangfkvgdvlireklnefa	132
Qy	142	TELTNQVISOYYNTEITVATQTMLDGHRVKRLDMTFAEGKPARVVDVINIIGNQHSFDADILI	201
Db	133	ksvkehvasvrynatvepivntlpnnraeilqineddkakiasltfkgnesvsstiq	192
Qy	202	DVLAIKONKINPLSKADRYTQEKLVTSLENIARAKYLNAGVPRFEIKADKLINEDKNRIF	261
Db	193	eqmlelqpswwkll-wgnkfegaqfekldqslrivylnngyakaqiktvdglndektkn	251
Qy	262	VEISLHEGEQYRFQTOFLGNLTVQAELEALLK-FKAEEGFSOAMLEQTTNNISFKFGD	320
Db	252	vtidvneqlqydlrsariignlgmsaeleplsalhldtfrsdiadvenaikaklge	311
Qy	321	DGYYYAQRVTRINDESRVTVDVEYYIDPVHPVYVRRINFTGNFKTODEVLRMRLEQES	380
Db	312	rgygsatvnsvpfdaddanktlaillvvdagrirtvqlrfegntvsadstlrqemrqeg	371
Qy	381	ALASNOXIOSRARELMTGFEKHHVTVDTR--PVPNSPDQVDVNFVVEEQSGSSTTAAGY	438
Db	372	twynsqivekgirldrtgfe--tvenridpingsndevdvvykcentgsinfqiy	429
Qy	439	SQSGGVTFQFDVSONNPMGTGKHHNASFSRSETREYVSLGTMNPYTFVNGVSQSLSGYYR	498
Db	430	gtesgisvqasvkdnlflgaaavsiagtkndygtvsnlgytepyfkkdgv--slggnvf	487
Qy	499	KTKYDNK---NISWYVLOSXGSLSGYGPIDENORISFGL-----NADNTKHLHGGRFMI	550
Db	488	fenydnksdtsnykrttygsnvtlqfpvnennsyvglghtynklsnfaleynrnlly	547
Qy	551	SNVKQLMADGKKIQVDNNGIPIPDFKHDTYTNAILGWNYSLSLDRPVPTQGMHSVD--LT	608
Db	548	qsmk-----fkgngiktndfdfs-----fgwnynslrnygtfkgvkaslggrvt	592
Qy	609	VGFGEKTHQKVYOGNTRYRP-----IKKSVLRYGAKLGYGN-NLPVFENFYAGGYGS	660
Db	593	ipgsdnkyykllsadvgqfypldrhlwvvsakasagyan-gfgnkrpbfyqytyagigls	651
Qy	661	VRGYDQSSLGRSQAYLTARRGOQTTLG-----EVMGYNALATGSELILPLPKPGDW	713
Db	652	lrgfaygysignpa---liyaevyngsgtgtfkkissdvignalataaeelivptpfvsk	708
Qy	714	I-DQVRVPIFTEGGQVFDTT-GMDKOTIDLTQPKDPOATAEQNAKAANRPLLTQDKOLRY	771
Db	709	sqntvrtslfvdaasvwnktwksdknglesdvk-----rlpdygksrira	755
Qy	772	SAGVGATWYTPIGLTSISYAKPLNKKONQDQTDTVQFQISVF	813
Db	756	stvgvfawqspqlpvlfsyvakplkkyvenddeqvfqfsiaqsf	797

Search completed: September 5, 2002, 09:54:02
Job time: 441 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:54:36 : Search time 37.39 seconds
(without alignments)
531.105 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 4202
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTDTVOFQIGSVF 813

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1052.5	25.0	797	3	US-08-433-522A-2
2	1052.5	25.0	797	3	US-08-433-522A-4
3	1052.5	25.0	797	3	US-08-433-522A-6
4	1052.5	25.0	797	3	US-09-135-166-2
5	1052.5	25.0	797	3	US-09-135-166-4
6	1052.5	25.0	797	3	US-09-135-166-6
7	1052.5	25.0	797	3	US-08-942-046-2
8	1052.5	25.0	797	4	US-08-942-046-4
9	1052.5	25.0	797	4	US-08-942-046-6
10	1047.5	24.9	793	3	US-08-433-522A-10
11	1047.5	24.9	793	3	US-09-135-166-10
12	1047.5	24.9	793	4	US-08-942-046-10
13	1042	24.8	792	3	US-08-433-522A-8
14	1042	24.8	792	3	US-08-433-522A-8
15	1042	24.8	792	4	US-08-942-046-8
16	146.5	3.5	2123	4	US-08-968-685A-10
17	142	3.4	2199	5	PCT-US95-11684-2
18	141	3.4	1599	2	US-08-617-697-9
19	138	3.3	1536	1	US-08-038-682-2
20	138	3.3	1536	1	US-08-302-832-2
21	138	3.3	1536	2	US-08-530-198-2
22	138	3.3	1536	2	US-08-469-880-2
23	138	3.3	1536	2	US-08-728-470-2
24	138	3.3	1536	2	US-08-617-697-2
25	138	3.3	1536	4	US-08-719-641-2
26	136.5	3.2	2314	4	US-08-268-347-49
27	136	3.2	990	4	US-09-627-376-7

28	134.5	3.2	1178	1	US-08-446-486-5	Sequence 5, Appli
29	134.5	3.2	1178	1	US-08-463-308-5	Sequence 5, Appli
30	134.5	3.2	1182	1	US-08-349-867-34	Sequence 34, Appli
31	134.5	3.2	1182	2	US-08-598-305A-34	Sequence 34, Appli
32	134.5	3.2	1188	1	US-08-239-476-34	Sequence 34, Appli
33	134.5	3.2	1188	2	US-08-639-923A-34	Sequence 34, Appli
34	134.5	3.2	1188	5	PCT-US95-05431-34	Sequence 34, Appli
35	134	3.2	2048	4	US-09-268-347-48	Sequence 48, Appli
36	132	3.1	905	4	US-09-074-658-70	Sequence 70, Appli
37	130	3.1	969	1	US-07-671-817A-4	Sequence 4, Appli
38	130	3.1	1177	1	US-07-828-788A-8	Sequence 8, Appli
39	130	3.1	1177	1	US-07-920-085-2	Sequence 2, Appli
40	130	3.1	1177	5	PCT-US92-11337-8	Sequence 8, Appli
41	130	3.1	1177	6	5169629-2	Patent No. 5169629
42	129.5	3.1	2089	1	US-08-418-893D-23	Sequence 23, Appli
43	129.5	3.1	2089	1	US-08-418-893D-24	Sequence 24, Appli
44	128.5	3.1	682	3	US-08-613-009A-10	Sequence 10, Appli
45	128.5	3.1	702	2	US-08-867-941-25	Sequence 25, Appli

ALIGNMENTS

RESULT 1
US-08-433-522A-2
; Sequence 2, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: STA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCE ADDRESSES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jib
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-433-522A-2

Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;
QY 23 STHAAQAFMNDITITGLQRTVIESLQSVLPRLQGVVSENQLADGVKALYATGNFSDV 82

```
Db 14 TTTFAAPFAVKDIKVDGVGDLQEOIRASLPVRAGQRTDNDVANIYRSLFVSGRFDV 73
Qy 83 QVYHOEGRI-IYQVTERPLIAEINPEGNRLPKGEOGLKNAGLAVGLOPKQATVOMIE 141
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVPTTEALKONLDANGFKVGDVLRKLNFEA 132
Qy 142 TELNQYISQGYNTFIVTKOTMLDGNRVKLDMTFAEGKPARVVDINIGNHQFSDADLI 201
Db 133 KSVKEHYASVGRYNATVEIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSTLQ 192
Qy 202 DVLAIKONKINPLSKADRYTQEKLVTSLENLRKYLNAGFVRFEIKDAKLINEDKNRIF 261
Db 193 EOMELQPSDWKWL-WGNKFEQAQFEKDLQSIYDYLNNYAKAQITKTDVQLNDEKTKVN 251
Qy 262 VEISLHEGEQYRFGQTOPLGNLTQOAELEALLK-FRAEEGFSOAMLEPQTNNISTKFGD 320
Db 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRSDIADVENAKALIGE 311
Qy 321 DGYYYAQIRPVTRINDESRVVDVEYIDPVHPVYVRRINFTGNFKTODEVLRREMRLQEG 380
Db 312 RYGSATVNSVPDFDDANKTLAITLVVDAGRRLTVRQLRFEQNTVSADSTLRQEMRQEG 371
Qy 381 ALASNOKIOLSRARLMRTGFEKHHVTVDTR--PVPNSPDQDVNFVVEQPSGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVKVKNRTGSIKFGY 429
Qy 439 SQSGGVTFQFDVSONNFMGTGKHYNASFSRSETREVYSLGMTNPFYTVNGVYSQSLGYR 498
Db 430 GTESSISYQASVKODNFGTGAAYSIAGTKNDYGTSVNLGYTEPYFKDGV--SLGNNVF 487
Qy 499 KTKYDNK---NISNYLDSYGSLSYGYPIDENQRISEGL-----NADNTKLHGGRPMGI 550
Db 488 FENYDNKSDTSSNYKRTYGSNTVLGFPVNNENNSYVGLGHTYKINSFALEYNRNLYI 547
Qy 551 SNVKOLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLLDRPVFPFTQGNHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYFPFTKGVKASLGGRVT 592
Qy 609 VFGDKTHQKVYQGNIRYRPF-----IKKSVLRGKYLGYGN-NLPFFYENFYAGGYGS 660
Db 593 IPGSDKNYKLSADVOGFYPLDRHLVWVSAKASAGYAN-GFGNKRLPFYQTYTAGGIGS 651
Qy 661 VRGYDQSSLPGRSAYLTARGQOITLG-----EVVGNALATFGESEILPLPFKGDW 713
Db 652 LRFGAYSGISGPN---IYAEYNGSGTGTFFKKTSSDVIIGNATATASAEELIVPTPFVSDK 708
Qy 714 I-DQVRPVIFIEGGQVFDTT-GMDKQITDLTQFKDPQATAEQNAKANRPLLTQDKOLRY 771
Db 709 SONTVRSLEFDAASVWNTKWSKXGLESVDLK-----RLPDYKSSKIRA 755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDTVQFOIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDDVEQFOFSGISGF 797
```

RESULT 2

```
US-08-433-522A-4
; Sequence 4, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
```

```
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/433,522A
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-4
```

```
Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;
Qy 23 STHAAADFWANDITITGLQRTVETIESLQSVLPFRIGVGVSENOADGVKALYATGNFSDV 82
Db 14 TTTFAAPFAVKDIRVQGVQDLEQQIRASLPVRAGQRTDNDVANIYRSLFVSGRFDV 73
Qy 83 QVYHOEGRI-IYQVTERPLIAEINPEGNRLPKGEOGLKNAGLAVGLOPKQATVOMIE 141
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVPTTEALKONLDANGFKVGDVLRKLNFEA 132
Qy 142 TELNQYISQGYNTFIVTKOTMLDGNRVKLDMTFAEGKPARVVDINIGNHQFSDADLI 201
Db 133 KSVKEHYASVGRYNATVEIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSTLQ 192
Qy 202 DVLAIKONKINPLSKADRYTQEKLVTSLENLRKYLNAGFVRFEIKDAKLINEDKNRIF 261
Db 193 EOMELQPSDWKWL-WGNKFEQAQFEKDLQSIYDYLNNYAKAQITKTDVQLNDEKTKVN 251
Qy 262 VEISLHEGEQYRFGQTOPLGNLTQOAELEALLK-FRAEEGFSOAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRSDIADVENAKALIGE 311
Qy 321 DGYYYAQIRPVTRINDESRVVDVEYIDPVHPVYVRRINFTGNFKTODEVLRREMRLQEG 380
Db 312 RYGSATVNSVPDFDDANKTLAITLVVDAGRRLTVRQLRFEQNTVSADSTLRQEMRQEG 371
Qy 381 ALASNOKIOLSRARLMRTGFEKHHVTVDTR--PVPNSPDQDVNFVVEQPSGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVKVKNRTGSIKFGY 429
Qy 439 SQSGGVTFQFDVSONNFMGTGKHYNASFSRSETREVYSLGMTNPFYTVNGVYSQSLGYR 498
Db 430 GTESSISYQASVKODNFGTGAAYSIAGTKNDYGTSVNLGYTEPYFKDGV--SLGNNVF 487
Qy 499 KTKYDNK---NISNYLDSYGSLSYGYPIDENQRISEGL-----NADNTKLHGGRPMGI 550
Db 488 FENYDNKSDTSSNYKRTYGSNTVLGFPVNNENNSYVGLGHTYKINSFALEYNRNLYI 547
Qy 551 SNVKOLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLLDRPVFPFTQGNHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYFPFTKGVKASLGGRVT 592
Qy 609 VFGDKTHQKVYQGNIRYRPF-----IKKSVLRGKYLGYGN-NLPFFYENFYAGGYGS 660
```

```
Db 593 IPGSDNKYKLSADYQGFPLDRDLHWVVSASAGYAN-GFGNKRLLPFYQTYTAGGIGS 651
Qy 661 VRGYDOSSLGRPSQAYLTARRQQTTLG-----EVVGGNALATFGSELILPLPKGDW 713
Db 652 LRGFAYSGIPNA---IYAEYNGSGTGFGFKISSDVIGGNAIATASAEIIVPTPFVSDK 708
Qy 714 I-DQVRPVIFIEGGQVFDIT-GMDKQTDIDLTKQKQATAEQNAKANRPLLTQDKQLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYKSKSRIRA 755
Qy 772 SAGVGTATWTPIGPLSISYAKPLNKKQNDQTDVTFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSIGGSF 797

RESULT 3
US-08-433-522A-6
; Sequence 6, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-433-522A-6

Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

Qy 23 STHAQAQAFMNDITITGLQRVTISLSQVLPFRGQVVSQNLADGVKALYATGNFSDV 82
Db 14 TTTVFAAPFAKDVIRVDGVDGQLEQIRASLPVAGQRVTDNDVANIVRSLEVSGRFDDV 73
Qy 83 QVYHOGRI-IQVTERPLIAINEFENPLIPKEGLQELKNAGLAVGQPLQATVQMIE 141
Db 74 KA-HOEGDVLVVSVAKSIISDVKIKGNSVIPTKALKQNLKANGPKVGDVLIREKLINEFA 132
```

```
Qy 142 TELTNOYISQYNTETITVKQTMLDGNRVRKLDMTFAEGKPARVVDINIIGNOHFSADLI 201
Db 133 KSVKEHVASVGRYNATVEPIVNTLPNNRAEILIQINEDDOKAKLASLTFKGNESVSSTLQ 192
Qy 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNLRKAYLNAGVFRFEIKADKININEDKNRIF 261
Db 193 EQMELQPDSSWKL-WGNKFEQAQFEDLOSIRDYILNNGYAKAQITKTDVQLNDEKTKVN 251
Qy 262 VEISLHEGQYRFGQTOFLGNLTYYTQAELEALK-FKAEEGFSQAMLEQTTNNISLTKFGD 320
Db 252 VVIDVNEGLOYDLRSARIIGNLGGMSAELEPLLSALHLNDFRSDIADVENAIKAKLGE 311
Qy 321 DGYIAQIRPVTRINDESRVDVEYIDPVHPVYVRRINFTNFKTQDEVLREMRQLEG 380
Db 312 RGVGSATVNSVPDQDANKTLAITLVVDAGRRLTVQLRFEQNTVSADSLRQEMRQEG 371
Qy 381 ALASNOKIQLSRARLMRTGFFKHVTVDTTR--PVPNSPDQVDVNVFVEEQSGSSTTAAGY 438
Db 372 TWYNSQLVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVDVYVYKVRNTGSGINEGIGY 429
Qy 439 SOSGGVTQFDYSONNFMGTGKHVNASFERSERETREYVSLGTMTPYETVNGVSOSLSGYR 498
Db 430 GTESGISYQASVKQDNFLGTGAASVAGTAKNDYGTSVNLGTYTEPYFTKQGV--SLGNVF 487
Qy 499 KTKYDNK--NISNYVLDYSGSLSYGYPIDENQRIISFGL-----NADNTKLHGGRFMGI 550
Db 488 FENYDNKSDTSSNYKRTTYGTSNVTGLGFPVNNENSYVGLGHTYNTKISNFALEYNRLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDKHDYTTYNAILGNWYSSLDPRVPTQGMHSVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYPTPKGVKASLAGRVT 592
Qy 609 VGFQDKTHOKVYVQGNIRPE-----TKKSVLRGVAKLGYCN-NLPFYENFYAGSYGS 660
Db 593 IFGSDNKYKLSADYQGFPLDRDLHWVVSASAGYAN-GFGNKRLLPFYQTYTAGGIGS 651
Qy 661 VRGYDOSSLGRPSQAYLTARRQQTTLG-----EVVGGNALATFGSELILPLPKGDW 713
Db 652 LRGFAYSGIPNA---IYAEYNGSGTGFGFKISSDVIGGNAIATASAEIIVPTPFVSDK 708
Qy 714 I-DQVRPVIFIEGGQVFDIT-GMDKQTDIDLTKQKQATAEQNAKANRPLLTQDKQLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYKSKSRIRA 755
Qy 772 SAGVGTATWTPIGPLSISYAKPLNKKQNDQTDVTFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSIGGSF 797

RESULT 4
US-09-135-166-2
; Sequence 2, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-135-166-2

```

Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

23	QY	STHAQADFPANDITITGLQRTVIESLQSVLPRLQGVVSVENQOLADGVKALYATQNSFSD 82
14	Db	TTTTVAAPFAVKADIRVDGQGLQEQIRASLPVRAGQRTDNDVANIVRSLFVSGRFDV 73
83	QY	QVTHQBGRI-IYQVTERPLIAEINPEGNRLIPKEGLOEGLKNAGLAVGPOKQATWQMT 141
74	Db	KA-HQEGDVLVSVVAKSIISDVKIGNSVITEALKQNLNDANGFKVGQDVLIREKLNFEA 132
142	QY	TELTNOIYSQYYNTEITVQOTMLDGNRYKVLDMTFAEGKPARVVDINIINGHFSDDADI 201
133	Db	KSVKEHYASVGRYNATVEPVTNLPNNRAEIIQINEDDKAKLASLTIPKGNESVSSLTQ 192
202	QY	DVLAIKDNKINPLSKADRTQEBKLVTSLENRAKYLINAGVFVEIKDAKLINEDKNRLF 261
193	Db	EQMELQPDSSMWKL-WGNKEFGAQFQKLDQSIHDYILNNGYAKAQITKTDVQLNDSKTKVN 251
262	QY	VEISLHEGEQYRGQTFQCNLTQYTOAELEALK-FKAEEGSQAMLEOTTNNISTKFGD 320
252	Db	VTIDVNEGLQYDLRSARIIGNLGGMSAELEPLLALHLNDTFRSDIADVENAIKAKLGE 311
321	QY	DGYYYAQIRPVTRINDESTRVVEYIIDPVHPVYVRINFTGNFKTQDVLRRMRQLEG 380
312	Db	RGYSATVNSVPDQDANKTLAITLVVDAGRRLTVQLRFEQNTVSADSTLRQEMRQOEG 371
381	QY	ALASNKIOLSRARMLARTGFFKHVTVDTR--PVPNSPDQVDNVFVVEEOPSSSTIAAGY 438
372	Db	TWYNSQLVELGKTRLDRTGFFE--TVENRIDPINGSNDEVDVYVKERTGTSINFGTGY 429
439	QY	SQSGGVTFQFQVSONNFMCTGKHVNASFQRSETREYISLGMTNXPYFTVNGVSQSLSGYR 498
430	Db	GTESGISYQASVKODNPLFGTGAAVSITAGTKNDYGTGVSNTGYTEPVTDKGV--SLGGNVF 487
499	QY	KTIYDNK--NISNYILDSVSGSLSGYPLIDENQRIISFGL-----NADNTKLHGGRFPCI 550
488	Db	FENYDNKSQDTSNRYKRTYGSNVTLGFPVNENNSYVGLGHTYKNISFALLEYNRNIYI 547
551	QY	SNYKQLMADGGKIQVDNNGIPQKHDYTYNATLGNWYSSLDPRFPVPTQCMGSHSDV--LT 608
548	Db	QSMK-----FKNGIKTNDFFS-----FGHNYNSLNRGYEPTKGVKASLGRVT 592
609	QY	VFGEDKTHOKVYVQGNIIYRPF-----TKKSVLRGYAKLGYGN-NLPFYENFYAGGYGS 660
593	Db	IPGSDNKYTKLSADVGQFPLDRHDLHWYVSAKASAGYAN-GFCNKRRLPYQYRTTAGGTS 651

```

QY 661 VRGVDQSSLGPRQAYLTARRGQOTTLG-----EWWGNALATGSELILPLPKGDW 713
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 652 LRGFAYGISGPN-----IYAEYGNMGSGTFTPKKISSVDVIGNATATASAEILVPTPFVSDK 708
QY 714 I-DQVRPVIIEGQVFDTT-GMDKQTDILTQFKDPQATAEQNAKAANRPLLTQDKOLRY 771
   : :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 709 SQNTVRTSLFVDAASVWNTKWKDKNGLSDVLK-----RLPDYKSSRIRA 755
QY 772 SAGVAGATWYPIGPILSISYAKPLNKKQNDQTDVQFOIGSVF 813
   I||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 756 STGVGFQWQSPIGPLFVSYAKPIKKYENDDDVEQFQSIGGSF 797
   :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 5
US-09-135-166-4
: Sequence 4, Application US/09135166
: Patent No. 6083743
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/135,166
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 797 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-135-166-4

```

Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

[illegible]


```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-135-166-6

Query Match          25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY      23 STHQAADFPHANDITITGLORVTITSELSQVLPPFLQGVSVSENQLADGKALYATGNFSDV 82
Db       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
14 TTTVFAPFPVAKDIRDVGVDGDEQQIRASLPVRAGQRTVDNDVANIVRSFLVSGREDDV 73
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      83 QVYHGEGR-I-YOYTERPLTAIEINFEGNRLIPKEGLEGLKNAGLAVGOPLKQATQVMIE 141
Db       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
74 KA-HQEGDLVVSVVAKSISDVYKIGNSVIPTALKQNLDANGFKVGDVLRLEKLNEFA 132
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY     142 TELTNOYISOGYYNTEITVTKTMLDGNRVLKIDMTFAEKGPKARVVDVINIIGNHFSDALI 201
Db       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
133 KSVKEHYASVGRYNATVEPTVNTLPNNRAETLIQINEDDKAKLASLTFFKGNESVSSSTLQ 192
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY     202 DVLAIKDNKINPLSKADRYTQEKLVTSLENLRARYLNAGFVRPEIKDAKLINEDKNRIF 261
Db       :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
193 EQMELQPDSNWKKL-WGNKFEGAQEKDLQSIRDYLLNNGYAKAQTITKDVLNDKERTKVN 251
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY     262 VEISLHEGEYRFQGTQFLGNLYTQAELEALLK-FRAEEGFQSALEQTTNNISTTFGD 320
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
           :| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Qy	321	DGYIAQIRPVTRINDRESRTDVEYYIDPVHPVVYVRINFETGNFKTQDEVLRMRMQLJEG	380
Db	312	RGYGSATVNSVPFDDANKTLAITLVVDAGRRLTVROLRFGENTVSADSTLRQEMRQOEG	371
Qy	381	ALASNQIKIUSRAKLMRTGFEKHHVTVDT--PVPNSPDQVDVNVFVEQSPGSGSTIAAGY	438
Db	372	TWNSQLVELGKILRDLRTGFFE--TVSNRIRDPINGSNDEVVVVKVERNTGSGINFGTGY	429
Qy	439	SQSGVTFQFDVSONNPMWCTGKHVNASFSESRSEPREVYSLGNTPNPYFTVNGVSQSLSGYYR	498
Db	430	GTEGSIYQASVKQDNFLGTGAAVSIAGTNDYGTSVNLGYTEPYTKDGV--SLGNGVYF	487
Qy	499	KTKYDNK---NISNYLDSYSGSLSYGYPIDENORISFGL-----NADNTKLGHGREFMI	550
Db	488	PENYDNSKDTSSNYKRTTGVSNVTGLFPVNNNSYVVGHTYNNKISNFALEYNRNDIY	547
Qy	551	SNYKQLMADGGKIQVDNNGGIPDEKFKHDYTTYNAILLGWNYSSLDPRVPYFTQGMHSVD--LT	608
Db	548	QSMK-----FKNGIKTNDPDSF-----FGWNYNSLRCGYFTFKGVKASLGRVY	592
Qy	609	VGFQDKTHQKVYQGNITRPF-----IKKSLVRGAKLYGYN-NLPFTYENFYAGYGS	660
Db	593	IPGSDNKYKILSADVOCGYPIIDRDHILVVWSAKASAGVYV-CFGNKRIKPFYOTYTAGTIGS	651


```
Db 74 KA-HOEGDVLVSVVAKSIIDVKIKGNSVIPTTEALKONLDANGFKVGDVLIIRKELNEFA 132
Qy 142 TELTNOYISOGYVPEITVTKOTMLDGNRVKLDMTFAEKPAPRVVDINILGNQHFSDADLI 201
Db 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAELIQUINEDDKAKLASLTFKGNESVSSSTLQ 192
Qy 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENIIRAKYLNAGVFRFEIKAKLAKLINEDKNRIF 261
Db 193 BOMELQPDSSWKL-WGNKFEQAQFEKDLQSIIRDYVYLNNGYAKAQTITKDVQLNDEKTKVN 251
Qy 262 VEISLHEGEQYRFGOTQFLGNLTQOAELEALLK-FRAEEGFSQAMLEQTTNNISTFEGD 320
Db 252 VTIDVNEGLQDLRSARIILGNLGSMAELPLLSALHNDLFRSDIADVENAIKALGE 311
Qy 321 DGYYYAQIRPVTRINDESRVDEYIIDPVHPVYVRINFTGNFKTODEVLRREMROLEG 380
Db 312 RGYGSATVNSVPDFFDANKTLAITVVDAGRRLLTVRQLRPEGNTVSADSTLRQEMRQEG 371
Qy 381 ALASNQIKLSRARLMRTGFFKHVTVDTR--PVPNSPDQVDVNFVVEQPSGGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVDVYVKERNFTGSGIFGY 429
Qy 439 SOSGGVTFQFVSONNPMGTGKHVNASFSRSETREVYSLGNTNPFYTVNGVSQSLSGYR 498
Db 430 GTEGSIYQASVKODNFGTGAAYSIAGTKNDYCTSVNLGTYEFTFKDGV--SLGNGVF 487
Qy 499 KTKYDNK---NISNYVLDYSGGSLSYGYPIDENQIRISFGL-----NADNTKLHGGRFMI 550
Db 488 FENYDNKSDTSNNYKRTTYGNSVNTLFGPPVNNENSYVGLGHTYKNISFALEYNRNLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLLDRPVFTQGNHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWYNSLNRGYFPTKGKASLGRVT 592
Qy 609 VGFGRDKTHQVYVYOGNIYRPF-----IKKSVLRGYAKLGYGN-NLPFYENFYAGGYGS 660
Db 593 IPGSDNKKYKLSADVQGFYPLDRHLVWVSASAGYAN-GFGNKRLLPFYQTYTAGGIGS 651
Qy 661 VRGYDQSSLGPRSQAYLTARRGOQTTIG-----EVVGNALATFGSELILPLPFKGDW 713
Db 652 LRGPAYGISGPN---IYAEYNGSGTGTFKKISSDVIGGNATASAEALIVPTPFVSDK 708
Qy 714 I-DQVRVIFTEGGQVDDTT-GMDKQITDLTKQKDPQATAEQNAKANRPLLTQDKOLRY 771
Db 709 SONTVTSFLVDAASVWNTKSKDNKLGESDLK-----RLPDYKGSRRIRA 755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKONDQTDTVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSIGGSF 797
```

RESULT 10

```
US-08-433-522A-10
; Sequence 10, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MTS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-10
```

```
Query Match 24.9%; Score 1047.5; DB 3; Length 793;
Best Local Similarity 31.7%; Pred. No. 1.1e-75;
Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;

Qy 23 STHAAQADFANDITITGLQVRVTIESLQSVLPFRLGQVSVENQADGVKALYATGNFSDV 82
Db 14 TTVFAAPPFKDILRVGVQGDLEQQIRASLPVVRAGORTDNDVANIVRSLSVSGREDDV 73
Qy 83 QVYHQEGR-IYQVTERPLIAEINFEGRNLPKPEGLKAGLVAGQPLKQATVQWIE 141
Db 74 KA-HOEGDVLVSVVAKSIIDVKIKGNSVIPTTEALKONLDANGFKVGDVLIIRKELNEFA 132
Qy 142 TELTNOYISOGYVPEITVTKOTMLDGNRVKLDMTFAEKPAPRVVDINILGNQHFSDADLI 201
Db 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAELIQUINEDDKAKLASLTFKGNESVSSSTLQ 192
Qy 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENIIRAKYLNAGVFRFEIKAKLAKLINEDKNRIF 261
Db 193 BOMELQPDSSWKL-WGNKFEQAQFEKDLQSIIRDYVYLNNGYAKAQTITKDVQLNDEKTKVN 251
Qy 262 VEISLHEGEQYRFGOTQFLGNLTQOAELEALLK-FRAEEGFSQAMLEQTTNNISTFEGD 320
Db 252 VTIDVNEGLQDLRSARIILGNLGSMAELPLLSALHNDLFRSDIADVENAIKALGE 311
Qy 321 DGYYYAQIRPVTRINDESRVDEYIIDPVHPVYVRINFTGNFKTODEVLRREMROLEG 380
Db 312 RGYGNTVNSVPDFFDANKTLAITVVDAGRRLLTVRQLRPEGNTVSADSTLRQEMRQEG 371
Qy 381 ALASNQIKLSRARLMRTGFFKHVTVDTR--PVPNSPDQVDVNFVVEQPSGGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVDVYVKERNFTGSGIFGY 429
Qy 439 SOSGGVTFQFVSONNPMGTGKHVNASFSRSETREVYSLGNTNPFYTVNGVSQSLSGYR 498
Db 430 GTEGSIYQASVKODNFGTGAAYSIAGTKNDYCTSVNLGTYEFTFKDGV--SLGNGVF 487
Qy 499 KTKYDNK---NISNYVLDYSGGSLSYGYPIDENQIRISFGL-----NADNTKLHGGRFMI 550
Db 488 FENYDNKSDTSNNYKRTTYGNSVNTLFGPPVNNENSYVGLGHTYKNISFALEYNRNLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLLDRPVFTQGNHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWYNSLNRGYFPTKGKASLGRVT 592
Qy 609 VGFGRDKTHQVYVYOGNIYRPF-----IKKSVLRGYAKLGYGN-NLPFYENFYAGGYGS 660
Db 593 IPGSDNKKYKLSADVQGFYPLDRHLVWVSASAGYAN-GFGNKRLLPFYQTYTAGGIGS 651
Qy 661 VRGYDQSSLGPRSQAYLTARRGOQTTIG-----EVVGNALATFGSELILPLPFKGDW 713
Db 652 LRGPAYGISGPN---IYAEYNGSGTGTFKKISSDVIGGNATASAEALIVPTPFVSDK 708
Qy 714 I-DQVRVIFTEGGQVDDTT-GMDKQITDLTKQKDPQATAEQNAKANRPLLTQDKOLRY 771
Db 709 SONTVTSFLVDAASVWNTKSKDNKLGESDLK-----RLPDYKGSRRIRA 755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKONDQTDTVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSIGGSF 797
```

Db 652 LRGFAYGSGPNA-----IYAEHNGCTFNKISSDVIGGNAITASAEILVPTFPVDSKQNT 708
Qy 717 VRPVIFIEGGQVFDTT-GMDKOTIDLTOFKDPOATAEQNAKAAANRPLLTQDKOLRYSGV 775
Db 709 VRTSLFVDAASVWNTKWKSDNGLSKVLKD-----LPDYGKSSRIIRASTGV 755
Qy 776 GATWYTPIGPLSISYAKPLNKKNDQDTQVFOIGSVF 813
Db 756 GFQWSPIGPLVFSYAKPIKKYENDDDVEQFQFSGGSF 793

RESULT 11
US-09-135-166-10
; Sequence 10, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-135-166-10

Query Match 24.9%; Score 1047.5; DB 3; Length 793;
Best Local Similarity 31.7%; Pred. No. 1.1e-75;
Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;
Qy 23 STHQAADFMANDTITGLQRVTIESLOSVLPRGLGVVSENQLAGVKALYATGNFSDV 82
Db 14 TTTVFAAPFPVKDVRVGVQGLDEQOIRASLPVRAGQRTDNDVNIYRSLFVSGREDDV 73
Qy 83 QVYHQEGRI-IYQYTERPLIAEINFEGNRLIPKEGLOGLNAGLAVGLOPLKQATVQME 141
Db 74 KA-HQEGDVLVVVAKSIISDVKIGNSVPTALKONLDANGFKVGDVLIREKLNFEA 132

Qy 142 TELTNOYISQYNTETITVKQTMLDGNRYKLDMTFAEGKPARVVDYINIIIGNHFSADALI 201
Db 133 KSVKEHYASVGRNATVEPIVNTLPNNRAEILQINEDDKAKLASLTFPKGNESVSSSTIQ 192
Qy 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYLINAGFVRFPEIKDAKLINEDKNRIF 261
Db 193 EQMELQPSDSWML-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTQVLDNDEKTKVN 251
Qy 262 VEISLHEGEBOYRFGOTQFLGNLTYYTOAEALLK-FKAEEGFQSOAMLEOTTNNISTKFGD 320
Db 252 VTIDVNEGLOYDLRSARIIGNLGMSAELEPLLSALHLNDTFRSDIADVENAIKALGE 311
Qy 321 DGYVYQAIRPVTRINDESRITVDVEYIIDPVHPVYVRRINFTGNFKTQDVLRRMRQLEG 380
Db 312 RGYNTTVNSVPDFDDANKTLAITFVWDAGRLTLRQLRFEGTNSADSTLRLQEMRQOEG 371
Qy 381 ALASNOKIOLSRARLMRTGFFKHTVDTNR--PVPNSPOVDVNFVVEEOPSSSITAAQY 438
Db 372 TWYNSQLVELGKIRLDRTGFFE--TVENRIDPINGSNDEVDVYVYKVKERTGTSINFGICY 429
Qy 439 SOSGGVTQFDVSONNFMTGKGVNASFSRSETREVYISLGMTNPYFTVNGVSOISLSGYR 498
Db 430 GTESGISYQTSIKQDNFLCTGAANVSIAGTKNDYGTSVNLGYTEPYETKDCV--SLGGNIF 487
Qy 499 KTKYDNK---NISNYVLDYSGSLSYGYPIDENQRISFGL-----NADNTKLHGGRFMGI 550
Db 488 FENYDNKSDTSSNYKRTTYGSNVTLPFPVNNNSYVGLGHTYKNKISNFALEYNNRLYI 547
Qy 551 SNVKQLMADGGKIQVDNNGIPDFKHDTYTYNAILGNWNTSSLDPRVPFPTQCMHSHVD--LT 608
Db 548 QSMK-----FKNGIKRTNDFDS-----FGWNTYNSLNRGYFPTKGVKASLGRVY 592
Qy 609 VGFGDKTHQVY--YQG-----NIYRPFKTSVLGRYAKLGYGN-NLPYENFYAGGYGS 660
Db 593 IPGSDNKYKLSADVQGFYPLDRHRWVVSASAKAGYAN-GGNKRLPYQIYTAGGIGS 651
Qy 661 VRGYDOSSLGRPSQAYLTARRQQT---TLGEVVGGNALATFGSELILPLPKGDMWI-DQ 716
Db 652 LRGFAYGSGPNA---IYAEHNGCTFNKISSDVIGGNAITASAEILVPTFPVDSKQNT 708
Qy 717 VRPVIFIEGGQVFDTT-GMDKOTIDLTOFKDPOATAEQNAKAAANRPLLTQDKOLRYSGV 775
Db 709 VRTSLFVDAASVWNTKWKSDNGLSKVLKD-----LPDYGKSSRIIRASTGV 755
Qy 776 GATWYTPIGPLSISYAKPLNKKNDQDTQVFOIGSVF 813
Db 756 GFQWSPIGPLVFSYAKPIKKYENDDDVEQFQFSGGSF 793

RESULT 12
US-08-942-046-10
; Sequence 10, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/942,046
;
; FILING DATE:
;
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER: 08/433,522
;
; FILING DATE: 12-SEP-1995
;
; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: STEWART, Michael I
;
; REGISTRATION NUMBER: 24,973
;
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (416) 595-1155
;
; TELEFAX: (416) 595-1163
;
; INFORMATION FOR SEQ ID NO: 10:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 793 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
; DS-08-942-046-10

```

Query Match 24.9%; Score 1047.5; DB 4; Length 793;
Best Local Similarity 31.7%; Pred. No. 1.1e-75;
Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;

QY	23	STHAQAADFMANDITITIGLQRTVITSLQSVLPRLGOVVSENOLADGVKALYATGNFSVD	82
Db	14	TTTTVAAPFPKDIRVDGVQGDLEQOIRASLPVRAGROQTDNDVNAVIRSLFVSGRFDV	73
QY	83	QVYHOEGRI-IYQVTERPLTAEFNEGNNRLIPKEGLOEGLKNAGLAVGLOPKAQTQVOMTE	141
Db	74	KA-HQEGDVLVSVVAKSIISDVKKIGNSVITPEALKQMLDANGFKVGDOVLIREKLINEFA	132
QY	142	TELTNOYISOGYYNTEITVTKQTMLDGNRVKLDMTFAEGRPAPRVIIINIGNHOFSADLI	201
Db	133	KSVKEHYASVGRYNATVEPVTNLPNNRAEILQIINEDDKAKLASLTAFKGNESVSSTLO	192
QY	202	DVLAIADNKNPLSKADRTQOEKLVTSLENRAKYINAGFVRFEIKADKALININEDKNRIF	261
Db	193	EQMELQDPSWKEL-WGNKFEQAGFEKDLQAIRDYIYLNNGYAKAQITKTDVQUNDKRTKVN	251
QY	262	VEISLHEGEQYRGQTFGLNLTYYOAELEALK-FKABEGFSQAMLEOTTNNISTKFGD	320
Db	252	VTIDVNEGQYDLRSARIINGLGMSAELEPLLSALHLNDTFRSDIADVENAIRAKLGE	311
QY	321	DGYYYAQIRPVTRINDESRVDVEYIIDPVHPVYVRINFTGNFKTQDEVLRREMRQLE	380
Db	312	RGYGNTTVNSVPDFFDANKTLAIFVVDVAGRRLTVRQLRFEQNTVSADSTLRQEMRQEG	371
QY	381	ALASNOKIOLSRARMLRTGFFKHVTVDTR--PVPNSPDQVDVFNVEVEOPSGSSTIAGY	438
Db	372	TWYNSQIVELGKIRLDRDTGFFE--TVENRIDPINGSNDEVVDVYVKERNITGSIINFGIY	429
QY	439	SQSGGVTFQFQVSNQNNFMGTGKHVNASFRRSETREYISLGMTNXPYTVNGVSQSLSGYR	498
Db	430	GTESGISYQTSIKQDNFLGTGAASVLAGTKDYGTSVNLGYTEPYTKDGV--SLGGNIF	487
QY	499	KTYKDNK--NISNYVDLSYSGSLSGYPIDENQRISFGL-----NADNTKLHGGRPMGI	550
Db	488	FENYDNKSQDTSNRYKARTYGSNVTGLGFPVNENNYSYVGLGHTYKNKISNFALEYNRNLYI	547
QY	551	SNVYKQMLADGGKIQVDNNGIPDKHDYTYNAILGNWYSSLDPRVPFPTQGMSSHVD--LT	608
Db	548	QSMK-----FKNGIKITNDFDS-----FGWNYNSLNRGVEPPTKGVKASLGGRV	592
QY	609	VFGDKTHOKV--YOG-----NIVRPFTKVLRGYAKLGYGN-NLPYFENFYAGGYGS	660
Db	593	IPGSDNNKYTKLSDVQGFYPLDRDRHWVYSAKASAGYAN-GFCNKRLLPQYPTTAGGIGS	651

	Qy	661	VRGVDQSLSGRPSQAYLTARRGOQT---TLGEVVGGNALATPGESEILPLPFKGDWI-DQ 7116
	Dd	652	LRGFRAYSIGPNA--IYAEHGNFTNKISSDVIIGNAITTASAGELIVPTFPFYSDKSQNT 708
	Qy	717	VRPVIFTEGGGVFDTT-GMDKQTTIDLTFQKDPOATAEQNAAKANRPLLTKDKOLRYSAGV 775
	Dd	709	VRTSLFVDAASVWMTKNKSKNGLESKLXD-----LPDYGKSSRIIRASTGV 755
	Qy	776	GATWTYTIGPLISYSYAKPLNNKQNDDPTVFQIGSVWF 813
	Dd	756	GFOWSQSDIGLPIVSFAKPPIKKYEINDVEOFSTIGSGF 793

```

RESULT 13
US-08-433-522A-8
; Sequence 8, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SiA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:j-b
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 792 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-8

```

Query Match	24.8%	Score 1042;	DB 3;	Length 792;
Best Local Similarity	31.4%	Pred. No. 3.1e-75;		
Matches 257;	Conservative 159;	Mismatches 335;	Indels 68;	Gaps 20;

Qy	23	STHAQAADFMANDITITGLQRVYIESLSQVLPFLRGVVYSENQIADGVKALXYATGNFSDV	82
		: : : : : : : : : : :	
Db	14	TTTVAAPFAKDIRVDGVQGDLEQIIRASLPVRAGORTDNDVANIYRSFLVSGFFDDV	73
Qy	83	QVYHQGRI-IYOVTRPLIAEINFGNRLIPKEGLOEGLKNAGIAGVOPLOKQATVQMTIE	141
		: : : : : : : : : : :	
Db	74	KA-HQEGDLVWSWAKSIIISDKIGNSIITPEALQKQNLGDANGFKVGDILIREKLINEFA	132
Qy	142	TELTNOYISQYYNTEITVQKTMLDGNRVKLDMTFAEGRPARVDINIICNGHFSADLI	201
		: : : : : : : : : :	
Db	133	QSVKEHYASGRYNAIVEIVNTLPNNRAEIIQIINEDDKAKLASITFGKNSGVSSSTIQ	192

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 5, 2002, 09:55:54 ; Search time 69.95 seconds
(without alignments)
1116.807 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 4202
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTDTQFQIGSVF 813

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1451.5	34.5	797	2 H83190	probable outer mem
2	1262	30.0	795	2 AC0129	probable surface a
3	1218.5	29.0	803	2 A80530	outer membrane pro
4	1210	28.8	810	2 A84742	hypothetical prote
5	1210	28.8	810	2 C90651	hypothetical prote
6	1210	28.8	810	2 C85502	hypothetical prote
7	1203	28.6	803	2 B82099	surface antigen VC
8	1172.5	27.9	784	2 E82731	outer membrane ant
9	1115	26.5	797	2 G81228	outer membrane pro
10	1111	26.4	797	2 D82000	outer membrane pro
11	1053.5	25.1	808	2 F84102	protective surface
12	1052.5	25.0	797	2 JC4078	protective surface
13	754	17.9	768	2 D71726	outer membrane pro
14	734	17.5	774	2 D97527	ompl protein presu
15	734	17.5	774	2 A82746	group 1 outer memb
16	727.5	17.3	769	2 F87486	outer membrane pro
17	697	16.6	768	2 B97725	outer membrane pro
18	680.5	16.2	781	2 A83355	outer membrane pro
19	641	15.3	617	2 H84957	hypothetical prote
20	568.5	13.5	778	2 C70412	outer membrane pro
21	515	12.3	739	2 A81430	outer membrane pro
22	434	10.3	916	2 G84601	protective surface
23	429	10.2	906	2 F71910	probable outer mem
24	397.5	9.5	792	2 H81693	outer membrane pro
25	396.5	9.4	792	2 B71539	probable omp85 ana
26	395	9.4	790	2 D86528	omp85 analog [imp
27	395	9.4	790	2 D72094	probable outer mem
28	356.5	8.5	853	2 A71339	outer membrane pro
29	325	7.7	821	2 B70199	outer membrane pro

ALIGNMENTS

RESULT 1

H83190
probable outer membrane protein PA3648 [imported] - Pseudomonas aeruginosa (strain PA
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83190
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: H83190
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-797 <STO>
A:Cross-references: GB:AE004784; GB:AE004784; PIDN:g9949799; PIDN:AAG07036.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3648
C:Superfamily: protective surface antigen D-15

Query Match	34.5%	Score	1451.5	DB	2	Length	797
Best Local Similarity	36.9%	Pred. No.	4.2e-76				
Matches	308	Conservative	163	Mismatches	297	Indels	67
Gaps	13						
Qy	7	KGFQVSAMTAMVMVMVSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVVSQNL	66				
Db	2	KRFLPAL-LSALMIAEVHAES--FTVSDIRVNGLQVSAGSVFAALPLNVGETIDQAL	58				
Qy	67	ADGVKALYATGNFSDVQVYHQEGRIIYQVTERPLIAEINFEGRNLTPKEGLQGLKNAGL	126				
Db	59	VOATRSLEFTGTFQDQLGRDGNVLVTVVVRPSISIEIEGNAKSKEDLLKLGKOSGL	118				
Qy	127	AVGOLKQATVOMIEITFNQYISQGYVNTETIVKQITMDLNGRVKLDMTFAECKPARVVD	186				
Db	119	AGEIFORATLEGRNELORQYVQAQRYSAEINAEVQPRNRVALKININQETVAAISH	178				
Qy	187	INIIGNQHSADLIDVLAIK-DNKINPLSKADRYTQEKLVTSLENLRAKYLNAGVRFVE	245				
Db	179	INVVGNTVSEEDLTDLFELKTTNWLSPFKNDKDYAREKLSGDLRLSYLLDQYINMD	238				
Qy	246	IKDAKININEDKNRIEVEISLHEGEQYRFGQTOFLGNLYTQAELEALLKFAEFGSQA	305				
Db	239	IASTQVTSITPDKKHVYITVINEGEKYYTRDVKLTGDLKVPEEEVKRLLLVQKQVFSRK	298				
Qy	306	MLEQNTNNISTKFGDGYVYAOIRPVTRINDESRTVDDVEYIDPVHPVYVRRINFTGNFK	365				
Db	299	VMTTSDLLITRLUGNEGYTFANVNGVPEAHDDDKTQVSVTFVDPGKRAYNRINFGNTK	358				
Qy	366	TQDEVLRREMRQLEGALASNQIKQLSRRLMRTGFFKFKHTVTDTRPVPNSPDQDVNFVVE	425				

30	271	6.4	833	2	AF2089	hypothetical prote
31	256.5	6.1	578	2	B64012	hypothetical prote
32	256	6.1	853	2	D70304	hypothetical prote
33	254.5	6.1	617	2	AF2897	conserved hypotet
34	254.5	6.1	641	2	H97672	hypothetical prote
35	251.5	6.0	676	2	AE2417	hypothetical prote
36	249	5.9	711	2	A72375	hypothetical prote
37	236.5	5.6	846	2	F75525	outer membrane pro
38	230.5	5.5	578	2	AE0428	probable exported
39	230	5.5	577	2	AD1055	probable exported
40	228	5.4	577	2	F86119	hypothetical prote
41	227	5.4	577	2	S56445	hypothetical prote
42	227	5.4	577	2	F91278	hypothetical prote
43	224.5	5.3	861	2	S77409	hypothetical prote
44	221.5	5.3	579	2	F83327	conserved hypotet
45	198	4.7	623	2	AI3488	outer membrane pro

```
Db 359 TEDEVLRREMRQEGGWASTYLLDQSKARLERLGYEKEVNVETPAVPGTDDQVDVNSYE 418
QY 426 EOPSGSTTAAGYSQSGQVTFQFDVSONPMGTGKHVNASFSESRSEPREVYSLGTMNPPYT 485
Db 419 EOPSGSITASVGAQAGLILGGSISQNNFLGTGNKVSIGLRSEYQTYRNFGEVDPPYT 478
QY 486 VNGVSOSLSGYVRKTKYD--NKNISNVLDSYGGSLSYGYPIDENQISFGLNADNTKLH 543
Db 479 VDGVSILGYNAFYKTDYDELDVDVASYSVNSLGAAGSICGPISEYSLRITLYGLSVQDQID 538
QY 544 GREFMGISNVKQLMADGGKIQDNNGIPDFKHDTYTYNAILGWNYSLSLDRPVFPFGQMSH 603
Db 539 TGRYT-VDEIYDFLKEG-----DN-----FTNFKASICWSESTLNGKVLATRGHSQ 584
QY 604 SYDL--TVGFGDKTHQKVYQNIYRPFTKKSVLGRYAKLGYGN-----NLPFYEVEFA 655
Db 585 SLTLETLPGSLDFKIDYRGQVAPLTDNTMRFHFLGYGDGYGSTERLPFTENYIA 644
QY 656 GYGVSRYGDSLSGRSQAYLTARRGQQTTL-----GEYVGGNALAT 698
Db 645 GGFNSVGFKDTLGRSTPSV-ARNPDGTPMKNOGPDCKGRYDTPDQDPEAFGGNIIIT 703
QY 699 FGSELILPFPKGDWIDQVRPFIETEGGVFTTGMKQOTIDLTQFKDPQATAEQNAKA 758
Db 704 GGAELFLPFPVKDQ-RQLRTVLFWDVGSTFDCTPTTTNCQGIK-----749
QY 759 NRPLTQDKOLRYSGAGVATWTPIGPLISYAKPLNKKQNDQTDVQFOIGSVF 813
Db 750 -----TDNLASSVGVGLTWITALGPLSLFSLATPIKPDNAETQVQFSLGQTF 797

RESULT 2
AC0129
probable surface antigen YP01052 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0129
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0129
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-795 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89894.1; PID:g15979119; GSPDB:GN00175
C:Genetics:
C:Gene: YP01052
C:Superfamily: protective surface antigen D-15
```

```
Query Match 30.0%; Score 1262; DB 2; Length 795;
Best Local Similarity 35.7%; Pred. No. 3.8e-65;
Matches 292; Conservative 152; Mismatches 329; Indels 44; Gaps 14;

QY 14 MTMAVMVMVSTHAQAD-EMANDITITGLQRTVIESLQSVLPRLGQVSVENQLADGVKA 72
Db 6 LLIASLLFGSATYVGADGVVNDIHFEGLQAVGAALLNMPVRVGDVTSDDDDIGKTIRA 65
QY 73 LYATGNSFDVQVYHQEGRILYQVTERPLIAIEINFEGRNLRIPKEGLQBLKNAGLGOPL 132
Db 66 LFATGNFEDVRVLRDGNLTIVQKERTIASITFGSGNKAVKEDMLKNLEASGVVRGEAL 125
QY 133 KOATVQMIETELNQYISQGYNYETITVTKOTMLDGNRKLDMTFAEGKPARVVDINIIGN 192
Db 126 DRTTISNIEKLEDIFYSVGKASVAVVTPLRNVRDLKLVTETGVSAKIQIINIVG 185
QY 193 QHFSADALIDLVAIKD-----NKNINPLSKADR-YTOEKLVTSLENRAKVLNAGVRFEI 246
Db 186 HSTTDELISRFQIRDEVPMWNV-----GDRYQKQKLAGDLLETLRSFYLDGRYARFNI 240
```

```
QY 247 KDAKLINEDKNRIFVEISLHGEBOYRFGQTQFLGNLTYTQAEALLKFAEEGFSQAM 306
Db 241 DSTQVSLTPDKKGIYVITINITEGPQFKLNSVSGLAGHQAEBKLTKIEPCELLFNGSK 300
QY 307 LBQTTNNISTKFGDDGYIAQIIRPVTRINDESRITVDVVEYIIDPVHPVYVRRINFTGNFKT 366
Db 301 VTRMEDDITKMLGRYGAYPRVVTQPEINDDKTVKLHINVDAGNRFYVYRHIFRFGNDTS 360
QY 367 QDEVLRREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNSPDQVDVNFVVEE 426
Db 361 KDSVLRREMRQMEGAWLGNQDVEACKERLNLRLGYETVDVETQRPVGAADLVDTYKYKE 420
QY 427 OPSGSTTAAGYSQSGQVTFQFDVSONPMGTGKHVNASFSESRSEPREVYSLGTMNPPYT 486
Db 421 RNTGSLNFGIGYGTSGSVSFQVQDDNMLGTGNTVGTNGTKNDYQTYAEFTLMDPYETV 480
QY 487 NGVSQSLGYRKTIDYDNKNISNVLDYSGSLSYGYPIDENQISFGLNADNTKLHGR 546
Db 481 DGVSLGGRIFYNDPKADNADLSGYNSYGADGTILGFFINENNSLRVGGVYVHNDL-SDM 539
QY 547 FMGISNVKQLMADGGKIQDNNGIPDFKHDTYTYNAILGWNYSLSLDRPVFPFGQMSHVD 606
Db 540 LPQVAMWRYLESVGERPGYD--GREGFTTDDFTLN--LGWTYNNLDRGFFFTSGVKSSV 595
QY 607 -LTVFGDKTHQKVYQNIYRPF--IKKSVLRGYAKLGYGN-----NLPFYEYFAGG 657
Db 596 TKIIVPGSDNEFYKVTFTSAYQPLNEDRSWLGLRGLGYGDGSGKEMPFYENFYAGG 655
QY 658 YGVSRYGDSLSGRSQAYLTARRGQQTTLGGEVVGGNALATFGSELILPFPKGD-WIDQ 716
Db 656 SSTVRGFRSNNIGPKA-AYYANGATVNTSDAVGGNAMAASIELITPTTFISEKYSNS 714
QY 717 VRPVIFIEGGVFTTGMKQOTIDLTQFKDPQATAEQNAKAANRPLTIDQDKRLYSAGVG 776
Db 715 VRTSIFIDSGVFWDTNW-----ENTAK-----TRAAGIPDYCKASNIRVSAGVA 758
QY 777 ATWTPIGPLISYAKPLNKKQNDQTDVQFOIGSVF 813
Db 759 LQWMSPLGLPVFSYAKPVKDYEGDKSEQFQFNIGKTW 795

RESULT 3
AB0530
outer membrane protein precursor yaer [imported] - Salmonella enterica subsp. enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB0530
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AB0530
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08682.1; PID:g16501505; GSPDB:GN00176
C:Genetics:
C:Gene: yaer
C:Superfamily: protective surface antigen D-15
```

```
Query Match 29.0%; Score 1218.5; DB 2; Length 803;
Best Local Similarity 33.9%; Pred. No. 1.2e-62;
Matches 285; Conservative 150; Mismatches 327; Indels 79; Gaps 15;

QY 14 MTMAVMVMVMS-----THAQAADFMANDITITGLQRTVIESLQSVLPRLGQVSVENQLA 67
Db 1 MAMKKLLIASLFLSSATYVGAEGFVVKDHFEGQLQAVGAALLSMFVPTGDTVNDDEIS 60
```

	QY	68	DGVKALATGNFSDVQYYHQEGRIIIVQYTERPLIAEINFEGRNLPKBEGLQBLKNAGLA	127
		:	: : : : :	
	Ddb	61	NTRIALFATGFNFEDVRVLBGNTLILVQVKERTPIASITFSGNKYSKDMDMLKQNLSEASGV	120
		:	: : : : :	
	QY	128	VGQPLKAQTVMQIETELTNQYISQGYNYTEIFVKOTMIDGNRVKLDMTFAEGCKPARVVDI	187
		:	: : : : :	
	Ddb	121	VGESLDRTTLSDIEKLEDFYISVGKYSASVAKVVTPLPRNRVDLKLVPQEGVSAKIQOI	180
		:	: : : : :	
	QY	188	NIIGHNQHPSDADLIDLAIKD-----NKINPLSKADR-YTQEKLVTSLENLRAKYLNA GF	241
		:	: : : : :	
	Ddb	181	NIVGHAFSTEELISHFQLDRPVMNVV-----GDRKYQOKLAGDLLETLSYYLDRGY	235
		:	: : : : :	
	QY	242	VRFELKDAKLNEKDNRITFEVISLHEGROYREFGOTFLNGLTITYTAOEALLKFKAEEG	301
		:	: : : : :	
	Ddb	236	ARENDISTOVSLTPDKKGIIYIVNITEGDQYKLSGVQVSGNLGASHAELEKLTKEPEGL	295
		:	: : : : :	
	QY	302	FEOAMLEOPTNNISFKFGDDGYAYIAIPRVTIRINDESRTVDVEYYIDPVHPHYVRRINFT	361
		:	: : : : :	
	Ddb	296	YNGTKVTKMEDDOI KLLGRYGAYPRVQSOPINDADKTVKLRVNVDAGNRFFYRKIRFE	355
		:	: : : : :	
	QY	362	GNFKTDQEVLRBEMBOLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNSPDQVDVN	421
		:	: : : : :	
	Ddb	356	GNDTSKDSVLRBEMKOMEGAUGSOLDVOGKERLARLGFEFVTDVTDTRVPSQPQVDVY	415
		:	: : : : :	
	QY	422	FVVEQPQSGSSTIAAGYSQGGVTGFQFVSONFNFMGTGKHVNASFSRSRETRYVSLGMTN	481
		:	: : : : :	
	Ddb	416	YVKERNTGSPNFGIGYGTESGSVFQAGVQODNWLTGYSVGINGTKNDYQYISELVTN	475
		:	: : : : :	
	QY	482	PFTVNGVSQSLSGYRYRTKKYNKMISNVLYDSYGGSLSYGYPIDENORISFGLNADNTK	541
		:	: : : : :	
	Ddb	476	PYFTVDGVS LGRIIFYNFEADDADLSDYTNKSQYDVTGLGPPINEYNTLRAGLGYVHNK	535
		:	: : : : :	
	QY	542	LHGGRFMGSNNVK-OLMADGGKIQVDNNGIPDFKHDTTYTNAILGWNTSSIDRRPVEPTQG	600
		:	: : : : :	
	Ddb	536	L-----SNMQPQIAMDRYLESGMDPDASPADAADFENN--YGWTYNKLDRGFYPTDG	585
		:	: : : : :	
	QY	601	--MSHSVLTUVFGDKTHQKVYVQGNIRYPFTK---KSVLRGYAKLGYGNL-----PYFE	651
		:	: : : : :	
	Ddb	586	SRVNLTKGVTIPGSNKEYYKVSLDATATVVIDNDHKWVVLGRTRWGYDGLGKGKEMPFE	645
		:	: : : : :	
	QY	652	NPYAGGYSVRGCDOSLSLGRSQAVLTARRGOOTI-----LGEWVGDN	694
		:	: : : : :	
	Ddb	646	NPYAGGSTVRGFSQNTIGPKA----VYKNGAHTSWDDDDYEDCTQESGCKSDDAVGDN	701
		:	: : : : :	
	QY	695	ALATFGESELILPLPKGD-WIDQVRPFVIEGGQVFDTP-GMDKQITDITLFKDPQATAE	752
		:	: : : : :	
	Ddb	702	AMAVASLEFITPTPEISKYANSVRTSFWDMGTVYDINWDSPSSAPSDVPDYSDEP----	756
		:	: : : : :	
	QY	753	ONAKAANRPLLTDQKOLRYSAVGATWTTPPGPLSIYAKPLNKKQNQDQTDTRVQIQISV	812
		:	: : : : :	
	Ddb	757	-----GNIRMSAGIALQWMSPLGPLVFSYAQPKPKYDGDGAEQFOQFNIGHT	802
		:	: : : : :	
	QY	813	F 813	
		:	: : : : :	
	Ddb	803	W 803	
		:	: : : : :	
	RESULT	4		
	A64742		hypothetical protein b0177 - Escherichia coli	
	C:Species:	Escherichia coli		
	C:C.Date:	12-Sep-1997	#sequence_revision 17-Sep-1997 #text_change 29-Sep-1999	
	C:Accession:	A64742		
	R:R:Blattner,	F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;		
	A.: Rose, D.J.; Mau, B.; Shao, Y.			
	S:Science	277, 1453-1462, 1997		
	A:Title:	The complete genome sequence of Escherichia coli K-12.		
	A:Reference number:	A64720; MUID:97426617		
	A:Accession:	A64742		
	A:Status:	nucleic acid sequence not shown; translation not shown		
	A:Molecule type:	DNA		
	A:Residues:	1-810 <BLAST>		

A,Cross-references: GB:AE000127; GB:U00096; NID:g1786370; PTDN:RAC73288.1; PTD:g178663
A:Experimental source: Strain K-12, substrain MG1655
C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;
Best Local Similarity 34.0%; pred. No. 3.9e-62;
Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

14	QY	MTAVMVMVMS-----THAAQADAFMANDITITIGLARVTIETSIQSVLPFRIGQGVVSENQIA	67
15	Db	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
16	Db	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
17	Db	1 MAMKLLLIASLFFSSATVYGAEGVFVKDIHFEGELQARVAVGAALLSNPVRTGTDVTDNEDIS	60
18	QY	DGVKALYATGNFSDQVYVHQBGRILIIYQTERPLIAEINFEGRNLIPKBCLOGLKKNAGLA	127
19	QY	: : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
20	Db	NTIRALFATGNFEDVRVLRDGDTLLVQVKERTPIASITFSGNKSVKDDMLKQNLKQLEASGV	120
21	QY	VGQPLKQATVQMIEITLTHQYISQGYNPIETIVKQTMILDGNVRKVLDMTFAEGKPARVVDI	187
22	Db	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
23	Db	121 VGESLDRITIAIDIEKLEDFYYSVGKYSASVAVVTPLPRNRVDLKLVPQEGVSAEQOI	180
24	QY	188 NIIGNHFSADLIDLVAIKD-----NKINPLSKADR-YTQBKLVTSLENLRAKYLNAGF	241
25	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
26	Db	181 NIVGNHAFITDELISHFQLRDEVRPMNVV-----GDRKYKOKLAGDLETLRSYLDRCY	235
27	QY	242 VRFEIKDAKLINEDKNRIFVEISLHEGROYRFGQTQFLGNLTYYTQAELEALLKFKABEG	301
28	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
29	Db	236 ARFNIDSTQVSLTPDKKGIYVTVNTEGDQYKLSGVEYSGNLAGHSAETEQUTKEPGBL	295
30	QY	302 FSOAMLEQTNNISKFGDDGYAYQIIRPVTRINDESRTVDVEYYIDPVHPVYVRIINFT	361
31	QY	: : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
32	Db	296 YNGTKVTYKMEDDIKLLGRGYATPRVOSMPEINDADKTVKURVNDAGNRFFYVRKIRPE	355
33	QY	362 GNFKTQDEVLRERMQLEGALASNOKIQLSRARLMRTGFFKHTVTDTRPVPNSPDQVDVN	421
34	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
35	Db	356 GNDTSKDAVLRERMQEGAWLGSDLVQDGKERLNRLGFEFETVDTQTVRPGSPDOVDV	415
36	QY	422 FVVEQSPGSSIIAGYSQSGGVTFQFVDSQNNFMGTGKHVNASFSSRSTRETVYSLGWTN	481
37	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
38	Db	416 YXVKERNTGSFNGFYGYGTESGVSPQAGVQDDNWILGTGYAVGINGKNDYQTYAELSVTN	475
39	QY	482 PYFTVNGVYSQSLSGYYRKYTKYDNKNKINSVLYDSYGGSLSYGYPIDENQRISEFGLNADNTK	541
40	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
41	Db	476 PYFTVDGVSGLGRLFYNDQFQADADLSYTKSYGTQVTLGPPINEYNSLRAGLGYVHNS	535
42	QY	542 LHGGRFMGISNVKQLMA-----DGKIQVDNNGIPDFKHDTVTYNAILGNWYSSSLDRPV	595
43	Db	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
44	Db	536 L-----SNMQPVAMWRYLYSNGEHPSTSDQNSPKTDDFTFN--YGVTYNKLDRCY	585
45	QY	596 FPTQG--MSHVSYDLTVGFGDKTHQKVYVYQGNITYRPF--IKKSVLRYGAKLGYGNL----	647
46	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
47	Db	586 FPTDGSRVNLTCKVTIPGSDNEYKVTLDATATYVPIDDDHKWVLGRTRWGYGDLGGKE	645
48	QY	648 -PFYENFYAGGYGVRGYDQSSLGPRSS-----QA-----YLTFARGQQTTL---GEV	690
49	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
50	Db	646 MPFYENFYAGGSSYTRGFQSNITIGPKAVYFPHQASNYDPDYDECATQDGAKDLCKSDDA	705
51	QY	691 VCGNALATFGSELILPLPFKGD-WLDOVRPVFIIEGGGVEDTGTGMDKQOTIDLTFQFKDQA	749
52	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
53	Db	706 VCGNMAVASLEFIPTPFISDKYANSVRTSFFWDMGTWVDNTNWDSSQSYGYPTSDP----	763
54	QY	750 TAEQNAKAANRPLLTDQKQRYSAGVGATWTPTIGPLSISYAKPLNKKNDQDTVQFQOI	809
55	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
56	Db	764 -----SNIRMSAGIALQWNSPLGLPVFSYAQPFPKKYDGDKAQEQFQNI	806
57	QY	810 GSVF 813	
58	QY	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
59	Db	807 GKTW 810	
60	Db	I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :	
61	RESULT	5	
62	C30651		
63	biochemical protein Ec50179 (imported) - Escherichia coli (strain 0157:H7)		

RESULTS

C90651

hypothetical protein ECs0179 [imported] - *Escherichia coli* (strain O157:H7, substrate

C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: C90651
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genome reference strain O157:H7
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C90651
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAB33602.1; PID:gl3359635; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs0179
C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;
Best Local Similarity 34.0%; Pred. No. 3.9e-62;
Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

QY 14 MTMAVMNMS-----THAAADFMANDITITGLQRTVIESLSQSVLPFRLGQVVSQNLA 67
DB 1 MAMKLLIASLLFSSATVYGAEGFVVKDIHFEGLRVAVGAALLSPVRTGDTVNDIS 60
QY 68 DGKVALYATGNFSDVOVYHQEGRIIVQVTERPLIAINFEGRNLPKEGLOGLKNAGLA 127
DB 61 NTRALFATGNFEDVRVLRDGTLLVQVAKERTPIASITFSGNKSVKDDMLKQNLASGVR 120
QY 128 VGQPLKQATVQMTETELTNOYISQGYNTETITVKQTMLDGNRVKLDMTFAEGKPARVVDI 187
DB 121 VGESLDRTTIADIEKLEDFYYSVGKYSASVAVVTPLPNRVVDLKLVFQEGVSAEIQOI 180
QY 188 NIIGNHFSADLIDVLAIKD-----NKINPLSKADR-YTQEKLVTSLENLRKAYLNAGF 241
DB 181 NIVGNHAFTTDELISHFQLRDEVPWNV-----GDRKYQKQKLAGDLETLSRYILDRGY 235
QY 242 VRPEIKDAKINEDKNRIFVELSLHEGQYRFGQTFGLNLTYYTQAELEALLKFAEEG 301
DB 236 ARNIDSTQVSLTPDKKGIYVTNITEGQYKLSGVEVSGNLAGHSAEIQLTKEPGL 295
QY 302 FSQAMLEQTTNNISTKFGDDGYGYAQRVTRINDSRVTDVEYYIDPVHPVYVRINF 361
DB 296 YNGTKVTKMEDDITKLLGRYGAYPRVQSMPEINDADKTVKLRVNDAGNRFYVKIRFE 355
QY 422 FVVEEQPSGSSTIAAGYSQSGVTFQDYSONNFMGTGKHVNASFSRSETREYVSLGTMN 481
DB 416 YKVKERTGTSFNGIGYTESGVSFOAQVQDNLGTGYAVGKNGKNDYQTYVAELSVTN 475
QY 482 PYFTVNGVSQSLSGYRKYTKYDNKNISNVLDSYSGSLSYGYPIDENQRISFGLNADNTK 541
DB 476 PYFTVDSGLGRFLFYNDFOADDADLSYTNKSYGTDVTLGFPINEYNSLRAGLVVHNS 535
QY 542 LHGGRFMGINSVKQLMA-----DGCKIQVDNNGIPDEKHYDITYNAILGNWYSILDRPV 595
DB 536 L-----SNMQPQVAMRWYLSMGEHPSTSDQNSFKTDDETFN--YGWTYNKILDRGY 585
QY 596 FPTQG--MHSVSDVLTGFGDKTHQVYVQNGIYRPF--IKKSVLRGYAKLGYGNL----- 647
DB 586 FPTDGSRVNLTKVTPGSDNEYKYVLTATVPIDDDDHKVWVLRGTRWNGYDGLGGKE 645
QY 648 -PPYENFYAGGVSRYGQSSIGLPRS-----QA-----YLTARRGQQTFL-----GEV 690
DB 646 MPYENFYAGGVSRYGQSSIGLPRS-----QA-----YLTARRGQQTFL-----GEV 690
QY 691 VGNALATGSELILPLPKGD-WIDQVRPIIEGSGVDFDTGMDKOTIDLTQFKDPOA 749

DB 706 VGNAMAVASLEFITTPTPFISDKYANSVRTSFDFWDMGTVMNDTNDSSQYSGVPDYSDB-- 763
QY 750 TAEQNAKAANRPLLTQDKOLRYSGAGVATWYTPIGPLTSISYAKPLNKKQNDQDTDTVQFOI 809
DB 764 -----SNIRMSAGIALOWMSPLGPLVFSYAQPFKKYDGDCAEQAQFQFNI 806
QY 810 GSVF 813
DB 807 GKTW 810

RESULT 6
C85502
hypothetical protein yaeF [imported] - Escherichia coli (strain O157:H7, substrain ED
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85502
R:Perina, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouasis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85502
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <STO>
A:Cross-references: GB:AE005174; NID:gl2512902; PIDN:AAG54479.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: yaeF
C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;
Best Local Similarity 34.0%; Pred. No. 3.9e-62;
Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

QY 14 MTMAVMNMS-----THAAADFMANDITITGLQRTVIESLSQSVLPFRLGQVVSQNLA 67
DB 1 MAMKLLIASLLFSSATVYGAEGFVVKDIHFEGLRVAVGAALLSPVRTGDTVNDIS 60
QY 68 DGKVALYATGNFSDVOVYHQEGRIIVQVTERPLIAINFEGRNLPKEGLOGLKNAGLA 127
DB 61 NTRALFATGNFEDVRVLRDGTLLVQVAKERTPIASITFSGNKSVKDDMLKQNLASGVR 120
QY 128 VGQPLKQATVQMTETELTNOYISQGYNTETITVKQTMLDGNRVKLDMTFAEGKPARVVDI 187
DB 121 VGESLDRTTIADIEKLEDFYYSVGKYSASVAVVTPLPNRVVDLKLVFQEGVSAEIQOI 180
QY 188 NIIGNHFSADLIDVLAIKD-----NKINPLSKADR-YTQEKLVTSLENLRKAYLNAGF 241
DB 181 NIVGNHAFTTDELISHFQLRDEVPWNV-----GDRKYQKQKLAGDLETLSRYILDRGY 235
QY 242 VRPEIKDAKINEDKNRIFVELSLHEGQYRFGQTFGLNLTYYTQAELEALLKFAEEG 301
DB 236 ARNIDSTQVSLTPDKKGIYVTNITEGQYKLSGVEVSGNLAGHSAEIQLTKEPGL 295
QY 302 FSQAMLEQTTNNISTKFGDDGYGYAQRVTRINDSRVTDVEYYIDPVHPVYVRINF 361
DB 296 YNGTKVTKMEDDITKLLGRYGAYPRVQSMPEINDADKTVKLRVNDAGNRFYVKIRFE 355
QY 422 FVVEEQPSGSSTIAAGYSQSGVTFQDYSONNFMGTGKHVNASFSRSETREYVSLGTMN 481
DB 416 YKVKERTGTSFNGIGYTESGVSFOAQVQDNLGTGYAVGKNGKNDYQTYVAELSVTN 475
QY 482 PYFTVNGVSQSLSGYRKYTKYDNKNISNVLDSYSGSLSYGYPIDENQRISFGLNADNTK 541
DB 476 PYFTVDSGLGRFLFYNDFOADDADLSYTNKSYGTDVTLGFPINEYNSLRAGLVVHNS 535

```
QY 542 LHGGFEMGSLNVKQLMA-----DGGKIQVDNNGIPDFKHDTYTTNAILGWNYSLLDRPV 595
Db 536 L-----SNMQPQVAMWRYLSMGEHPSTSDQDNSEFTDDTFN--YGTWYNKLDRCY 585
QY 596 RPTQG--MHSVVDLTVGGDKTHQVYVYOGNLYRPF--IKKSVLRGYAKLGYGNL----647
Db 596 PFDGSRNLTOKVTPGSDNEYYKVTLDATYVPIDDDHKKWVVLGRTRWGYDGLGKRE 645
QY 648 -PFYENFYAGGYSVGRGVDQSSSLGPRS-----QA-----YLTARRQQQTLL---GBV 690
Db 646 MPFYENFYAGGSSTVRGQSNITGPKAVYFPHQASNYDPDYDECATQDQKDLCKSDDA 705
QY 691 VCGNALATFGSELIILPLFPKGD-WIDQVRPVITIEGGQVFDTGMDKQIDTIDTQKDPQA 749
Db 706 VCGNAMAVASLEFITPTFFISDKYANSVRTSFFWMGTVMWDNWDSSQSYGPDYSDP--763
QY 750 TAEQNAKAANRLLTQDKOLRYSAGVATWYTPIGPLSTSYAKPLNKKQNDQTDVQFOI 809
Db 764 -----SNIRMSAGIALQWMSPLGLPVFSYAQPPKKYDGDRAEQFQFNI 806
QY 810 GSVF 813
Db 807 GKTW 810

RESULT 7
BB2099
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: BB2099
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
A:Accession: BB2099
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <HE>
A:Cross-references: GB:AE004297; GB:AE003852; NID:g9656810; PIDN:AAF95396.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2252
A:Map position: 1
C:Superfamily: protective surface antigen D-15

Query Match 28.6%; Score 1203; DB 2; Length 803;
Best Local Similarity 33.4%; Pred. NO. 9.8e-62;
Matches 278; Conservative 155; Mismatches 328; Indels 72; Gaps 15;

QY 16 MAVVMVMTAAQAAD-EMANDITITGLORVTIESLQSVLPFRLGQVVSQENLADGVKALY 74
Db 8 LATLLTSVANGAEKFFVQDIIQDLQRLVGAALLMKVPRVGSQDVANIKALY 67
QY 75 ATGNFSDVQVYHQEGRIVYQVTERPLIAEINFEGRNLIPKEGLQBLKNAGLAVGQPLKQ 134
Db 68 SSGNFEDVKVLRDGNLWVQKERTIASVSFSGNKAKEEQKLNLEASSIRVGEALDR 127
QY 135 ATVQMIEITELNQTYSQGYNTFETVTKOTMLDGNRVKLDMTFAEGRKPARVDINIGNOH 194
Db 128 TTLSNIERGLEDFYYSVGKNATVAVVTPPLPRNRADLKFVFTGEGVSAKIQIINFIGNOV 187
QY 195 FSDADLIDVIAKDKNI--NPLSKADRYTQEKLVTSLENLRKYLNAGFVREIKADKLN 252
Db 188 FSDEELLSRFLNVDVAMNFWLAD--DKYKKQVLGADIEALRTYYLDRGYLKFDVSTQYA 246
QY 253 INEDKNRIFVETSLHEGQYRGTQFLGNLTYYTQAELEALKFKAEGFSQAMLEOTTN 312
Db 247 ISPDKKGVYITLNLNHEGEPYTSKVQFRGELMGKAEFTSLIPFEIGETINGSVATRLKE 306
```

```
QY 313 NISTKFGDDGYAAQIRPVTRINDESRTVDVEYYIDPVHPVYVRINFTNFKTQDEVLR 372
Db 307 SVKKVLGSEAGYAPQVOTIPEFDEKQOQSLVVHVEAGKRVYVRDTRFVGNNSTRDEVLR 366
QY 373 REMROLEGALASNOKLTQLSRARLMRTGFFKHVTVDRPVPNSPDQVDVNFVBEQPSGSS 432
Db 367 REMROHEGSLNKSOKLETGKTRLNRLGFFETVEVQVVRVPGSEDDQVDLVTVSKEANSNV 426
QY 433 TTAAGYSQGGVTFFQFDVSONNFMGTGKHVNASFRSSETREVYSLGNTNPFYVNGVSQS 492
Db 427 NEGVGYGTESGVSFQVGLQDNFNLGSGNRVGVNAMINDYQKNLTLEYRDPYWNLDGVSIG 486
QY 493 LSGYRKTYDKNNKISNYVLDYSGLSYGYPIDENQISFGLNADNTKLHGFRFMGINS 552
Db 487 GKVFYNOFEASEAGIVDYTNESYGTSLTWGFPFDELNRFEFGIGYTHNKI--GNLTPYLQ 544
QY 553 VKOLMA-----DGGKIQVDNNGIPDFKHDTYTTNAILGWNYSLLDRPVFPPTQGHSHS- 604
Db 545 VENFLAAQASNIDSGGNLLTDD-----FDINLSWTNRNLSNYSFETAG-NHQ 591
QY 605 --VDLTVGFCDKTHQKVYQGNIRPFIKKS-----VLRGYAKLGYGN--LP 648
Db 592 AFYKMTVPGSDAQYFKLQYDVRYQYFPLTKKHEFTLLRG--RLGYNGYQOTDGNLFP 649
QY 649 FYENFYAGGYSVGRYDQSSSLGPRS--QAYLTARRGOQTTLGEVVGGNALATFGESELILP 706
Db 650 FYENFYAGGFTSLRGFGNSAGPKAVYRDYSGSNNGSDTATDDSVGSGNALALASVELIVP 709
QY 707 LFPKGDWI--DQVRPVIFIEGGQVFDITGMDKQID----LTFQKQPOATAEQNAKANR 760
Db 710 TPFASEARNQRTISFYDMASVMDFEYRGKADYGNQYVYDSDP-----756
QY 761 PLLTDQKQLRYSAGVGATWYTPIGPLSTSYAKPLNKKQNDQTDVTOFOIGSVF 813
Db 757 -----TNRSSYGVALQWVSPMGPLVSLAKPIKKEGDEEFFFTTIGRTF 803

RESULT 8
E82731
outer membrane antigen XFI046 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: E82731
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82731
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-784 <SIM>
A:Cross-references: GB:AE003941; GB:AE003849; NID:g9105978; PIDN:AAF83856.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fra
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuranae, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, C.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XFI046
C:Superfamily: protective surface antigen D-15
```

[illegible]

RESULT 9
G81228
outer membrane protein Omp85 NMB0182 [Imported] - Neisseria meningitidis (strain MC58 serotype 4) [NCBI]
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81228
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Neilson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Rhee, H.; Olin, K.; Vamathevan, J.; Gill, J.; Scarfaro, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venter, A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755
A:Accession: G81228
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-797 <TET>
A:Cross-references: GB:AE002375; GB:AE002098; NID:g7225394; PIDN:AAF40639.1; PID:g7225394
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0182
C:Superfamily: protective surface antigen D-15

Query Match	26.5%	Score 1115;	DB 2;	Length 797;
Best Local Similarity	32.5%	Pred. No. 1.2e-56;		
Matches 266;	Conservative 154;	Mismatches 361;	Indels 38;	Gaps 14;
Qy	10	QVSWTAVMVMVMTSHAAQADFMANDITIGLQRTVIESLQSVLPFRLGQVVSENQLADG	69	
Db	2	KLKQIASALMLGISPALADFTTQDIRVEGLQRTSTVFNYLPVKVGVDYNDTHGSAI	61	
Qy	70	VKALYATGNFSDVOYHQEGRITIIQVTERPIATAEINFGNRLIEPGEQLQEKGNAGLAVG	129	
Db	62	IKSLYATGFFDVRVETADGQLLTIVTERPTGISLNTGAKMLQNDALKKNLESGLAQS	121	
Qy	130	QPLQAIVQMIETELTNQYISQGYNTEITVKQTMDCGNRVKLDMTFAEGKPARVDINI	189	
Db	122	QYFQWALNQAVGLKEEYLGKGLNTQITPKVTKLARNRVGIDITIDECKSAKITIEF	181	
Qy	190	IGNGHFSDADLIDVLAIKDNKI-NPLSKADRYTQEKLVTSLENARAKYLNAGFVRFEEKD	248	
Db	182	EGNQVYSDRKLMRQMSLTEGGIWTWLTFRSNQFNQKQFAQDMEKYTFYQNNGYDFRILD	241	
Qy	249	AKLINEDEKNRIFVEISLHEGEQYRFQGTQPLGNLT-YTQAEALBALKFKFAEGFQSAML	307	
Db	242	TDIQTNEDKTKQTIKITVHEGGRFWGKVSIEGDTNEVPKAELEKLLTMKPGKYERQOM	301	
Qy	308	EQTTNNISTKFGDDGYAAQIRPVTRINDSESTVDVEYIDPVHPVVRRTNFTNGKTO	367	
Db	302	TAVLGETQNRMSGAGYAYSLSVQPLNAETKTVDVLHIEPGRKIIYNEIHTGNKNTR	361	
Qy	368	DEVLRMRQLEGALASNQKIQLSRALMRGTGPKHKVHTVDTRPVPSPDQVDVNFVVEEQ	427	
Db	362	DEVYRRELROMESAPYDTSKLQRSKERVELLIGFDYDNQVDAVPLAGTPDKVDLNMSLTER	421	
Qy	428	PSGSTTAAGYSQSGGVTFFQDVSONFMGTGKHVNASFRRSETREYVSLGCMNTPYETVN	487	
Db	422	STGSLDLSAGWQDTGLVMGSAVSODNLFTGKSAALRASRKTTLNGSLSTFDPIYETAD	481	
Qy	488	GVSOSSLGYR--KTKYDNKNISNYLDSYGGSLSYGYPIDENQRIISFGLNADNTKLHG	545	
Db	482	GVSLGYDVGKAFDPRKASTSIQYKTTAGAGIRMSVPVTEYDRVNFGLVAEHLTVN--	539	
Qy	546	RFWGISNVKQLMADGGKIQVDNNGCIPDEKHDYTTYNAILGNWYSSLDPRFPPTQGHSHV	605	
Db	540	--TYNKAPHYADFIKKYKGTGD-TDGSFKGLYKCTGVGMGRNKTDLSALWRTGYLTGV	595	
Qy	606	DLTVGF-GDK-THOKVYVOGNIYRPFITKKSVLR-----GYAKLGYG--NNLPFVFNFYAG	656	
Db	596	NAEIALPGSKLQYYSATHNQTWFPPLSKTFTMLGGEVGIAG-GYGRTEKIEIPFENYGG	654	
Qy	657	YGSVRGYDOSSLGPRSQAYLTARGOQTLTGEVV--GGNALATFGSELILPLPFKGDWI	714	
Db	655	GLGSVRGYESTGLGPK-----VYDEYGEKISYGGNKNKANVSAELLFPMPGAKD-A	703	
Qy	715	DQVRPVIFIEGGGVDFPTGMDKOTIDLTFQDKPQATQEAQNAANRPLLTQDKQLRYSAG	774	
Db	704	RTVRLSLFADAGSVDGKTYYDNNSSAT-----GGRVQNIYGAGNTHKSTFTTNELRYSAG	758	
Qy	775	VGATWYTPIGPLSISYAKPLNKKONDQTDTPVQFQISVF	813	
Db	759	GAYTWISPLGPMKFSYAYPLKKRPEDEIQRFQFOLGTTF	797	

RESULT 9
G81228
outer membrane protein Omp85 NMB0182 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_accession: G81228
R:Retelton, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleisig, O.; Olin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Fraser, C.M. Science 287, 1809-1815, 2000
A:Title: Complete genome sequence of Neisseria meningitidis

Query Match	25.1%	Score 1053.5	DB 2	Length 808
Best Local Similarity	31.8%	Pred. No. 4.4e-53		
Matches	260	Conservative 156	Mismatches 337	Indels 63
Gaps	20			
Qy	23	STHAQAADFAMNDITITGLQRVTTESLQSVLPFRIGOVSVSNOLADGKALYATGNFSDV	82	
Db	27	TTTVAAPFAKDIRVGVQGGDLFQQIRASLRPVRAGQRVTDNDVANIVRSFLVSGRFDVV	86	
Qy	83	QVYHGEGRIRIYQVTERPLPIAEINFEGRNLIPKEGLKAGLAVGQPKQATVOMIE	141	
Db	87	KA-HOEGDGLVSVVAKSIISDVKIKGSIITPEALKQNLNDANGFKVGDVLIREKLNEFA	145	
Qy	142	TELTNOYISQGYNYTEITVKOTMLDGNRVKLDMTFAEGKPARVVDIINIQNQHSADLI	201	
Db	146	KSVKEHYASVGRYNAVTEPIVNTLPNNRAETLIQINEDDDRAKLASLTFKGNGSVSSLTQ	205	
Qy	202	DVLAIKDNKINPLSKADRYTOEKLVTSLENLRARYKLNAGFVFRIKDAKLINEDKKNRIF	261	
Db	206	BOMELQDPSWKLL-WGNKFECAQFEKDLQISRDYLLNNGYAKAOKITTDVOLNDEKTKVN	264	
Qy	262	VEISLHEGEYRFQGTQFPLGNLTTQBAELEALLK-FRAEEGFSQAMLEQTTNNITSTKFGD	320	
Db	265	VTIDVNEGLODLRSARIIGNLGGMSAELEPLLALSALHNDTFRRSADIADVENAIKALGE	324	
Qy	321	DGYVYAIQRPYTRINDESRTVDVEYIDPVHPVVVRINFTNGFKTODEVLRREMRQLEG	380	
Db	325	RGYGSATVNSVPDDEDRANKTLAITLVDAGRRLTVRQLRFGNTVSADSTLURQEOMQEG	384	
Qy	381	ALASNOKTOLSRALMRITGFPHVTVDTNR--PVPNSDPDQDVNFVWEEQPSGSGSTIAAGY	438	
Db	385	TWYNSQLVELGKIRLDRGTGFE--TVENRIDPINGSNDEVDVYKVKERTGTSINFGIGY	442	
Qy	439	SQSGGVTFQFDVSONNFMGTCKHYNASFSRSTREYVSLGMTNPYFTVNGVSQSLSGYYR	498	
Db	443	GTESGISYQASVKQDNFLGTGAASVIAGTKNDYCTSVNLGTYTEPYFTKDGV--SLGGNVF	500	
Qy	499	KTKYDNK---NISNVLDISYGSLSYGPIDENQRISEGL-----NADNTKLHGCRPMGI	550	
Db	501	FENYDNKSDTSFSSNKRITYTSNTLTGFPVNNENSSYYVGLGHTYKNKISNFALFYNNRLYI	560	
Qy	551	SNVKOLMADGGKIOVDNNGIPDEFKHDTTYTNAILGWNYSSLDPRVFTPTQGHSHSDV--LT	608	

Db 561 QSMK-----FKNGIKTNDPDS-----FKGNSVLRGKAGLGYGN-NLPFFYENFYAGGVRV 605
QY 609 VFGDKTHQKVYQGNIRYRP-----IKKSVLRGKAGLGYGN-NLPFFYENFYAGGVRV 660
Db 606 IPGSDNKKYKLSADVQGFYPLDRDLHWVVSASAKAGYAN-GFGNKRPLFFYTYTAGGIGS 664
QY 661 VRGYDQSSGLSPRSQAYLTARRGQOT---TLGEVVGNALATFGSELILPLPFKGDWI-DQ 716
Db 665 LRGFAYGSGIPGNA-IYAEHNGNGCTFKKISSDVIGGNAITATASAEELIVPTPFVSDKQNT 723
QY 717 VRPIETIEGGQVDTT-GMDKQITDLTQFKDPQATQONAKAANRPLLTQDKQLRYAGV 775
Db 724 VRTSLFVDAASVWNTWKWSKSLGD-----NNVLKSLPDYGGKSSRIASTGV 770
QY 776 GATWYTPIGPLSISYAKPLNKKQNDQDTDTVOFGISVF 813
Db 771 GFQWQSPIGPLVFSYAKPIKKYENDVQFQFSGISGF 808
RESULT 12
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
A:Variety: type b
C:Accession: JC4078
C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
R:Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus inf
A:Reference number: JC4078; MUID:95255676
A:Accession: JC4078
A:Molecule type: DNA
A:Residues: 1-797 <FLA>
A:Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448
A:Experimental source: type b
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 25.0%; Score 1052.5; DB 2; Length 797;
Best Local Similarity 31.8%; Pred. No. 4.9e-53;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;
QY 23 THAQAAADFMANDITITGLQRTVIESLQSVLPFRLGQVSENQADGVKALYATGNFSDV 82
Db 14 TTTVFAAPEVAKDIRVDGVQDLEQQIRASLPVRAGQRTDNDVANIVRSLSFGREDDV 73
QY 83 QVTHQEGRI-IYQVTERPLIAEFNENRLLPKEGLOGLKNAGLAVGQPLKQATVQMI 141
Db 74 KA-HOEGDVLVSVVAKSIISDVKIGNSVIPTEALKONLDANGFKVGDVLIKREKLENEA 132
QY 142 TELTNQYISOGYKNTETVKTQMLDGNRKLDTFAEGKPARVVDINIGNQHFSADLI 201
Db 133 KSKVHVASGRNATVEPIVNTLPNNRAEILIQINEDDKAKLASITFRKGNSSVSTLQ 192
QY 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENLRKAKYINAGVFRFEIKDAKININEDKNRIF 261
Db 193 EQMELQDPSWML-WGNKPEGAQFEDQLQSDIRYLLNNGYAKAQITKTDVQLNDEKTKVN 251
QY 262 VELSLHEGEQYRGQTFQFLGNLYTQAELEALK-KFAEGFGQAMLEQTTNISTKFGD 320
Db 252 VTIDVNEGLOYDLRSARIIGNLGMASAELEPLLSALHLNDFRSDIADVENAIAKALGE 311
QY 321 DGYYYAQIRVTRINDESRRVDVEYIDPVHVVYVRINFTGNFKTQDEVLREMRQLEG 380
Db 312 RYGSATVNSVPDFDANKTLATLVVDAGRRLTVKQLRPEGTVVSADSTLRQMRQEG 371
QY 381 ALASNKIQLSRARLMRTGFFKHVTVDTR--PVPNSPDQVDVNFVVEEQPSGSSITAAQ 438
Db 372 TWYNSQVLELGRDLRTGFFE--TVENRIDPINGSNDEVVYVKERTGTSINFGICY 429

QY 439 SQSGCVTFQEDVSONFMGTGKHVNASFSRSETREVYSLGMTNPYFTVNGYSQSLSGYR 498
Db 430 GTESSISYQASVKQDNFLGTGAAYSIAGTKNDYGTSVNLGYTEPYFTKDGV--SLGNNVF 487
QY 499 KTKYDNK---NISNVLDYSYSGSLSYGYPIDENQIRISFGL-----NADNTKLHGGRPMGI 550
Db 488 FENYDNSKSDTSSNYKRTTYGTSNVTGLFPVNNNSYVGLGHTYNTKISNFALEYNRNLYI 547
QY 551 SNVQQLMADGGKIQVDNNGIPDPFKHDYTYNAILGWYSSLDPRVFPPTQGHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDPDS-----FGWYNSLNRGYFTFGVKASLGGVRV 592
QY 609 VFGDKTHQKVYQGNIRYRP-----IKKSVLRGKAGLGYGN-NLPFFYENFYAGGVRV 660
Db 593 IPGSDNKKYKLSADVQGFYPLDRDLHWVVSASAKAGYAN-GFGNKRPLFFYTYTAGGIGS 651
QY 661 VRGYDQSSGLSPRSQAYLTARRGQOTTLG-----EVVGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGSGIPGNA-IYAEYNGSGTGTFFKISSDVIGGNAIATASAEELIVPTPFVSDK 708
QY 714 I-DOVRPVIETIEGGQVDTT-GMDKQITDLTQFKDPQATQONAKAANRPLLTQDKQLRY 771
Db 709 SQNTVRTSLFVDAASVWNTWKWSKNGLESVDLK-----RLPDYGGKSSRIRA 755
QY 772 SAGYGATWYTPIGPLSISYAKPLNKKQNDQDTDTVOFGISVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVQFQFSGISGF 797
RESULT 13
D71726
outer membrane protein omp1 (omp1) Rp160 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: D71726
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichert-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: D71726
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-768 <AND>
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14627.1; PID:g386
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: omp1; Rp160
C:Superfamily: protective surface antigen D-15

Query Match 17.9%; Score 754; DB 2; Length 768;
Best Local Similarity 26.3%; Pred. No. 8.4e-36;
Matches 216; Conservative 145; Mismatches 383; Indels 76; Gaps 20;
QY 7 KGFOVSAMTMAVMVMSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSENQL 66
Db 2 KIISKILTKILLTIFYHISFADYVRKITIEGNHVRSTIESYLKLVNGETYNNKS 61
QY 67 ADGVKALYATGNSDQVY-HQEGRIYQVTERPLIAEFNENRLLPKEGLOGLKNAG 125
Db 62 DEAIKRLIATSLPRNIMYITNDGNLIVNVTETPTFSSVVFSGNSIKTKTILAKEITYMS 121
QY 126 LAVGQPLKQATVQMIETELTNQYISOGYKNTETVKTQMLDGNRKLDTFAEGKPARV 185
Db 122 ---GESLSQAKIELDKVKKILEIYKSRGRFSTKTPKLSLENNRVKIVFDIAEGPKTVIK 178
QY 186 DINIIGNQHSADLIDVLAIKONK-INPLSKADRYTQEKLVTSLENLRKAKYINAGVFR 244
Db 179 SIYFSGNEHYSDSELKSILVTKESRWFRLNESNDYDPRVEYDKELLREFYQSVGADF 238
QY 245 EIKDAKLINEDKNRIFVELSLHEGEQYRGQTFQFLGNLYTQ-AELEALLKFAEGSGS 303

Db 239 RYISASVALNDKTEVTTYSIEEGEKYRFGNVTTIDNKLNTINIKOLAKNIVIKOGKIFN 298
Qy 304 QAMLEQTTNNISTKFGDDGYIAQRPVTRINDESRITVDVEYIDPVHPVYVRRNETGN 363
Db 299 MKTVDDIAEKIGEYFANGPAVNYPIKND-NHTADIFIEBKADKVINIKINIINN 357
Qy 364 FKTODEVLRRMRQLEGALASNOQTOLSRARLMRTGFFKHVTVDRPVPNSDQVDVNFV 423
Db 358 LKTEDHVIRRAFKIBEGDVNRSYTEKGERNLRLNDYFEKVSISLAQT-KAKDKYDVNVE 416
Qy 424 YEEOPSGSSTIAAGYSQSGVTFQDVSONNFMGTGKHVNASFSRSEYVSLGWTNYP 483
Db 417 VDEKSTSIGDLGNTAGGLFGRFSLERNLVGTGKLLNAGVQVSKNSTSYGCGITDHP 476
Qy 484 FTWNGVSQSLGGY----RKTNDKNKNSINYVLDYSGSLSGYGPIDE--NORISFGLNA 537
Db 477 FLDRDLSLVNAFRNYTRGASVLTNTDQSYKLSHIGKISLGYDMKEDLSHEDILIKR 536
Qy 538 DNTKLHGRFRFGISNVKOLMADGGKIQVDNNGIPDFKHDTYTYNAILGNWYSSSLDRPVP 597
Db 537 DILSAPSP-----SNSIFLNEQMGKLTISAIG-----HTIT-----YDQTDNKIVP 577
Qy 598 TQG--MHSVDLTGCFGDKTHQVYVYOGNIYRPFIKKSVLRCYAKLG-----YGNLPLF 649
Db 578 KNGLVSTQTFAGVGGDNKYIKHIECKFYKSFNNKVTKLKSAAGDMAGLGKMWRI 637
Qy 650 YENFYAGGYGVGRYDQSSLSGRSQAYLTARRGQOTTILGEVVGYNALATFGSELILPLF 709
Db 638 SDRFNIGDY-SLRFPASGVGR-----EKNTEGLGGERYYTFTSTELNFPVP 685
Qy 710 KGDWIDQVRPVIFIEGGQVFTTGMDKOTIDLTPKDPQATAEQNAKAAANRPLLTQDKOL 769
Db 686 PEEF--NFTGAVFIDLGSVMG-VGLNKK-----QYKTPNG-----FYNDQSL 724
Qy 770 RYSAGVGATWTPIGPLSISYAKPLNKKQNDQTDVQFQI 809
Db 725 RASVGFGIWTRFAPIRMDMGFFPKKKQYD--DFQNEHL 762

RESULT 14
D97527
ompl protein precursor (U515683) [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: D97527
A: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
A.: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: D97527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87173.1; PID:g15156447; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2554
A:Map position: circular chromosome

Query Match 17.5%; Score 734; DB 2; Length 774;
Best Local Similarity 26.4%; Pred. No. 1.2e-34;
Matches 219; Conservative 150; Mismatches 354; Indels 106; Gaps 22;

Qy 15 TMAYMVMTSHAQADFNANDITITGLORVTIESLQSVLPRLGOVSVSENLADGVKALY 74
Db 22 SVAGLGLVASAGVANAAYISIDVRGASGADSVNRITAPGNKFNNSDIDESVKRLY 81
Qy 75 ATGNFSDQVYHQEGRIITYQVTRPLIAEINFEGRNLPLPKGLQBLGNAGLAVGQPL-- 132
Db 82 ATGFYSNVMSRVSGSTLVVTVNENQVNVFNGNRKKDD-----KLAVIQVQTPMGP 135
Qy 133 -KQATVQMIEFTLNQYISQGYNTETITVKTMLDGNRVKLDMTFAEGKPARVVDINIIG 191

Db 136 FNQAVTADIAIRKEAYSAGRSDEVTITQTVSVGGRVNIAPVINEGERTKIGRIDFIG 195
Qy 192 NOHFSADALIDLVL-AIKDNKINPLSKADRYTQEKLVTSLENLRAKYLNAGVFRFEKDAK 250
Db 196 NNSYSDGRALAAVINTKKSNNLSFLTRKOVYNEDKURADEALRQFYNNRGYADFRVYSD 255
Qy 251 LMINEDKNRIFVEISLHEGEQYRFGQTQFLGNLTYYQ-AEALALKFAKEEGFSQAMLEQ 309
Db 256 AVLDESKENEYTISIVDEGKKYDFGNVAVESTVPGVDSGSELOGLVETROGASYSAREVOQ 315
Qy 310 TTNNISTKFGDDGYIAQRPVTRINDESRITVDVEYIDPVHPVYVRRNETGNFTQDE 369
Db 316 SMEAISKRVRAGEYFPFARVTPRGDRDMSGNTIGVTYIVDQGERAYVERIEIRGNTTRDY 375
Qy 370 VLRRMRQLEGALASNOQTOLSRARLMRTGFFKHVTVDRPVPNSDQVDVNFVVEQPS 429
Db 376 VIRREDISEGDAFNQTTITAAKRRLEALGYFSKVNISTAG-GSAPDRVVIVVDEDQST 434
Qy 430 GSSTIAAGYSQSGVTFQDVSONNFMGTGKHVN-ASFRRSETREYVSLGWTNYPYFTVNG 488
Db 435 GSEFGAGYSQNDGVLLEASVEEKNFLGRQYIRVAAGAGEDDARTYLSLFTPEYF--- 490
Qy 489 VSQSLSGYRKYTKYDNKNISNTVLDYSGSLSGYGPIDENQRISFGLNADNTKLHGGRFM 548
Db 491 LGYRLAAGFDLEFKNOSKSEYNYDEQGFALRVTAITENLSTFKY----- 537
Qy 549 GISNVKQLMADGGKIQVDNNGIPDFKHDTYTYNAIL--GW-----NYSSSLDRPVF 596
Db 538 ----TYKQINVEGKGWNNANLAE-----PYQALIRGEDWTQSILSNLTNLTNVDLRNM 588
Qy 597 PTQGMHSHVDLTVGF----GDKTHQKVYVQGNLYRPFIKK----SVLRGYA--KLGYGNN 646
Db 589 PREGWQA--LTNEFAGLGGDSEYKYIYAKARYYTLSDYDVGISLTQAGHVMFTGDN 646
Qy 647 LPFYENFYAGGYGVGRYDQSSLSGRSQAYLTARRGQOTTILG-EVVGGNALATFGSELIL 705
Db 647 LLVDFQFQFGG-RQVRGFKNDGIGPR-----IGSDSIGGTTTYFAASAEEVTA 691
Qy 706 PUP-FKGWDIDQVRPVIFIEGGQVFTTGMDKOTIDLTPKDPQATAEQNAKAAANRPLLT 764
Db 692 PMPGVEDP--GLRAGFVDACTMGKNVSTQTV-----KD----- 726
Qy 765 QDKQLRYSAGVGATWTPIGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 727 -DNSIRASAGIGVMASPPGPIRVDAIPIAKEDYDEEQRFREGMSNTF 774

RESULT 15
AE2746
group 1 outer membrane protein precursor ompl [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AE2746
A:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Fam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE2746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <KUR>
A:Cross-references: GB:AE008688; PIDN:AA42387.1; PID:g17739796; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: ompl
A:Map position: circular chromosome

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:03:36 ; Search time 36.65 Seconds
(without alignments)
858,908 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 4202
Sequence: 1 MRNSYKFGQVSAMTAVMM.....LNKKQNDQDTVQFIGSVF 813

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	1210	28.8	810	1	UP05_ECOLI	P39170 escherichia
2	1053.5	25.1	795	1	D152_HAEIN	P44935 haemophilus
3	1052.5	25.0	793	1	D153_HAEIN	O32629 haemophilus
4	1052.5	25.0	797	1	D151_HAEIN	P46024 haemophilus
5	641	15.3	617	1	Y237_BUCAI	P57331 buchnera ap
6	256.5	6.1	578	1	YTFM_HAEIN	P44038 haemophilus
7	227	5.4	577	1	YTFM_ECOLI	P39320 escherichia
8	197.5	4.7	469	1	CS51_HUMAN	O9Y512 homo sapien
9	171	4.1	463	1	U140_DROME	O9V784 drosophila
10	157.5	3.7	1577	1	HLYA_PROMI	P16466 proteus mir
11	154.5	3.7	1256	1	MRP_STRSU	P32653 streptococc
12	149	3.5	1902	1	P3P_LACLC	P15292 lactococcus
13	148	3.5	475	1	YDF6_SCHPO	O10478 schizosach
14	146	3.5	826	1	YEBB_ECOLI	P33341 escherichia
15	146	3.5	1902	1	P1P_LACLC	P16271 lactococcus
16	143	3.4	1902	1	P2P_LACPA	O02470 lactobacill
17	142.5	3.4	705	1	PPCF_FLAME	P27195 flavobacter
18	140.5	3.3	1902	1	P2P_LACLC	P15293 lactococcus
19	140	3.3	2201	1	TENA_HUMAN	P24821 homo sapien
20	139.5	3.3	1324	1	CUT3_SCHPO	P41004 schizosach
21	137.5	3.3	735	1	FCT_ERWCH	Q47162 erwinia chr
22	136	3.2	917	1	HXA3_HAEIN	P54355 haemophilus
23	135.5	3.2	1332	1	XKDO_BACSU	P54334 bacillus su
24	134.5	3.2	866	1	YCBS_ECOLI	P75857 escherichia
25	134.5	3.2	1178	1	YJAC_BACTK	P05068 bacillus th
26	134	3.2	655	1	YKDA_MYCCA	P45615 mycoplasma
27	133.5	3.2	937	1	CS32_ECOLI	P15484 escherichia
28	133	3.2	398	1	YLB1_CAEEL	P46576 caenorhabdi
29	133	3.2	1048	1	AGOL_ARATH	O04379 arabidopsis
30	132.5	3.2	863	1	Y8TO_ECOLI	P33924 escherichia
31	130.5	3.1	1957	1	YD86_SCHPO	Q10411 schizosach
32	128.5	3.1	905	1	HXAL_HAEIN	P44602 haemophilus
33	127.5	3.0	839	1	YDDB_HAEIN	P45182 haemophilus

34	127	3.0	484	1	YNG6_YEAST	P53969 saccharomyc
35	127	3.0	922	1	DPOI_RICFE	O9raa9 rickettsia
36	126.5	3.0	1258	1	ICEN_ERWHE	P16239 erwinia her
37	126	3.0	925	1	VPH_BHPH1	P51735 bacterioph
38	126	3.0	1034	1	ICEN_PANAN	Q47879 pantoea ana
39	125.5	3.0	880	1	LYTD_BACSU	P39848 bacillus su
40	125.5	3.0	2022	1	ANT1_ONCVO	P21249 onchocerca
41	124.5	3.0	1341	1	VG37_BPT2	P07067 bacterioph
42	124.5	3.0	1754	1	PMPB_CHLTR	O84418 chlamydia t
43	124.5	3.0	2366	1	TOXB_CLODI	P18177 clostridium
44	124	3.0	1630	1	MSPI_PLAFK	P04932 plasmodium
45	124	3.0	1639	1	MSPI_PLAFW	P04933 plasmodium

ALIGNMENTS

RESULT 1

UP05_ECOLI

ID UP05_ECOLI STANDARD; PRT; 810 AA.

AC P39170: P39181: P77465;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 16-OCT-2001 (Rel. 40, Last annotation update)

DE Unknown protein from 2D-page spots M62/M63/O3/O9/T35 precursor.

GN YAEI OR B0177 OR Z0188 OR ECS0179.

OS Escherichia coli, and

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562, 83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RA "The complete genome sequence of Escherichia coli K-12.";

RT Science 277:1453-1474(1997).

RL [2]

RP SEQUENCE FROM N.A.

RC Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E., Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D., Davis R.W.;

RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

RL [3]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamowski K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RT Nature 409:529-533(2001).

RL [4]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RA "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.";

RT DNA Res. 8:11-22(2001).

RL [5]

RP SEQUENCE OF 21-32 AND 351-362.


```

Query Match      25.1%; Score 1053.5; DB 1; Length 795;
Best Local Similarity 31.8%; Pred. No. 1.9e-54;
Matches 260; Conservative 158; Mismatches 337; Indels 63; Gaps 20;

QY 23 STHAAQAFMAMDIITIGLQRTVIESLQSVLPFRIGQVVSSENOLADGVKALYATGNFSDV 82
DB 14 TTTTFAAPFAVKDIRVGVQGDLEQOIRASLPVRAQRVTDNDVANIIVRSFLVSGRFDV 73

QY 83 QVYHOGRI-IQVTERPLIAEINFEGRNLIPKEGLOEQLKAGLAVGLOPKQATVQMI 141
DB 74 KA-HOEGDVLVSVVAKSIISDVKIGNSIIPTEALKQNLKDFGKGVGLIREKLENEFA 132

QY 142 TELTNOYISQYINTEITVQTMDCNRVKLDMTFAEGKPARVVDINIIGNHESDADLI 201
DB 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSLTQ 192

QY 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYLNAGFVRPEIKDAKLINEDKNRIF 261
DB 193 EQMELQPDSSWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251

QY 262 VEISLHEGEQYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISKFGD 320
DB 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRRSDIADVENAIAKALGE 311

QY 321 DGYVYVAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRRMRQLEG 380
DB 312 RGYGNTVNSVPDFDDANKTLAITFVVDAGRRLTVQRLFEQNTVSADSTLRQMRQEG 371

QY 381 ALASNQIKQLSRARLMRTGTFKHTVTDTR--PVPNSPDQVDVNFVVEEOPSGSTIAAGY 438
DB 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVYKERTNGTSINFGICY 429

QY 439 SOSGGVTQFQVDSQNNFMCTGKHVNASPSRSETREVYSLGTMNPNYPTVNGVSOGLSGYR 498
DB 430 GTESGISYQTSIKQDNFLGTGAASVIACTKNDYGTGTVNLGTYTPYFTKDGV--SLGGINIF 487

QY 499 KTKYDNK---NLSNVLDYSGSLSYGYPIDENQISFGL-----NADNTKLHGGRFMGI 550
DB 488 FENYDNKSDTSSNYKRTTYGNSVTLGFPVNNENNSYVGLGHTYINKISNFALEYNRNLYI 547

Query Match      25.0%; Score 1052.5; DB 1; Length 793;
Best Local Similarity 31.8%; Pred. No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 65; Gaps 21;

QY 23 STHAAQAFMAMDIITIGLQRTVIESLQSVLPFRIGQVVSSENOLADGVKALYATGNFSDV 82
DB 14 TTTTFAAPFAVKDIRVGVQGDLEQOIRASLPVRAQRVTDNDVANIIVRSFLVSGRFDV 73

QY 83 QVYHOGRI-IQVTERPLIAEINFEGRNLIPKEGLOEQLKAGLAVGLOPKQATVQMI 141
DB 74 KA-HOEGDVLVSVVAKSIISDVKIGNSIIPTEALKQNLKDFGKGVGLIREKLENEFA 132

QY 142 TELTNOYISQYINTEITVQTMDCNRVKLDMTFAEGKPARVVDINIIGNHESDADLI 201
DB 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSLTQ 192

QY 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYLNAGFVRPEIKDAKLINEDKNRIF 261
DB 193 EQMELQPDSSWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251

QY 262 VEISLHEGEQYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISKFGD 320
DB 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRRSDIADVENAIAKALGE 311

QY 321 DGYVYVAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRRMRQLEG 380
DB 312 RGYGNTVNSVPDFDDANKTLAITFVVDAGRRLTVQRLFEQNTVSADSTLRQMRQEG 371

QY 381 ALASNQIKQLSRARLMRTGTFKHTVTDTR--PVPNSPDQVDVNFVVEEOPSGSTIAAGY 438
DB 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVYKERTNGTSINFGICY 429

QY 439 SOSGGVTQFQVDSQNNFMCTGKHVNASPSRSETREVYSLGTMNPNYPTVNGVSOGLSGYR 498
DB 430 GTESGISYQTSIKQDNFLGTGAASVIACTKNDYGTGTVNLGTYTPYFTKDGV--SLGGINIF 487

QY 499 KTKYDNK---NLSNVLDYSGSLSYGYPIDENQISFGL-----NADNTKLHGGRFMGI 550
DB 488 FENYDNKSDTSSNYKRTTYGNSVTLGFPVNNENNSYVGLGHTYINKISNFALEYNRNLYI 547

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
NCBI_TaxID=727;
[1]
SEQUENCE FROM N.A.
RC STRAIN=PAK 12085;
RA MEDLINE=97427952; PubMed=9284140;
RX Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U60834; AAB61977.1; -.
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surfaceAg; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFB2036801A1A CRC64;

Query Match      25.0%; Score 1052.5; DB 1; Length 793;
Best Local Similarity 31.8%; Pred. No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 65; Gaps 21;

QY 23 STHAAQAFMAMDIITIGLQRTVIESLQSVLPFRIGQVVSSENOLADGVKALYATGNFSDV 82
DB 14 TTTTFAAPFAVKDIRVGVQGDLEQOIRASLPVRAQRVTDNDVANIIVRSFLVSGRFDV 73

QY 83 QVYHOGRI-IQVTERPLIAEINFEGRNLIPKEGLOEQLKAGLAVGLOPKQATVQMI 141
DB 74 KA-HOEGDVLVSVVAKSIISDVKIGNSIIPTEALKQNLKDFGKGVGLIREKLENEFA 132

QY 142 TELTNOYISQYINTEITVQTMDCNRVKLDMTFAEGKPARVVDINIIGNHESDADLI 201
DB 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSLTQ 192

QY 202 DVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYLNAGFVRPEIKDAKLINEDKNRIF 261
DB 193 EQMELQPDSSWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251

QY 262 VEISLHEGEQYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISKFGD 320
DB 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDFRRSDIADVENAIAKALGE 311

QY 321 DGYVYVAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRRMRQLEG 380
DB 312 RGYGNTVNSVPDFDDANKTLAITFVVDAGRRLTVQRLFEQNTVSADSTLRQMRQEG 371

QY 381 ALASNQIKQLSRARLMRTGTFKHTVTDTR--PVPNSPDQVDVNFVVEEOPSGSTIAAGY 438
DB 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVYKERTNGTSINFGICY 429

QY 439 SOSGGVTQFQVDSQNNFMCTGKHVNASPSRSETREVYSLGTMNPNYPTVNGVSOGLSGYR 498
DB 430 GTESGISYQTSIKQDNFLGTGAASVIACTKNDYGTGTVNLGTYTPYFTKDGV--SLGGINIF 487

QY 499 KTKYDNK---NLSNVLDYSGSLSYGYPIDENQISFGL-----NADNTKLHGGRFMGI 550
DB 488 FENYDNKSDTSSNYKRTTYGNSVTLGFPVNNENNSYVGLGHTYINKISNFALEYNRNLYI 547

RESULT 3
D153_HAEIN STANDARD; PRT; 793 AA.
AC O32629;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
```

```
QY 551 SNVKQLMADGGKIQVDNNGIPDPKHDYTYNAILGNWYSSLDREVFPFQGMHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYFPTKGVKASLGRVT 592
QY 609 VGFQDKTHQV--YQG-----NIYRPFIKKSVLRGVAKLGYGN-NLPFYENFVAGGYGS 660
Db 593 IPGSDNKKYKLSADVQGFYPLDRHRWVVSASAGYAN-GFGNKRLPFFQTYTAGGIGS 651
QY 661 VRGYDQSSLSGRSQAYLTARRQOT--TLGEVVGGNALATFGSELILPLPFFKGDWT-DQ 716
Db 652 LRGFAYGSIGPNA---IVAEGHNGTENKISSDVIGGNATITASAEILVPPFVSKDSQNT 708
QY 717 VRPVIFTEGGGVFTT-GMDKQTDLTQFDQPOATAEQNAKAANRPLLTDQKQLRYSGV 775
Db 709 VRTSLFVDAASVWMTKWSKDGLESKVLKD-----LPDYGKSSRIRASTGV 755
QY 776 GATWYTPIGPLISISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDVDEQFQFSIGGSF 793

RESULT 4
D151_HAEIN STANDARD; PRT; 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Ag) (Outer membrane protein D15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT The sequencing of the 80-kDa D15 protective surface antigen of
RL Haemophilus influenzae.";
RN Gene 156:97-99(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=MINNA, AND EAGAN;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; U13961; AAA85645.1; -
DR EMBL; U60832; AAB61974.1; -
DR EMBL; U60833; AAB61976.1; -
DR InterPro; IPR000184; Bac_surAg_D15.
DR Pfam; PF01103; Bac_surfaceAg; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF1B CRC64;
```

```
Query Match 25.0% Score 1052.5 DB 1; Length 797;
Best Local Similarity 31.6% Pred No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY 23 STHAAADFMANDITITGLQRVTIESLSQSVLFRGLGVVSENOLADGVKALYATGNFSVD 82
Db 14 TTTVFAAPFAKDIRVDGVQDLEQQIRASLPVRAGQRTVDVANIVRSLSFVSGREDDV 73
QY 83 QVYHQEGRI-IYQTERPLIAEINFEGNRLIPKEGLQBLKAGLAVGQPLKQATVQMIE 141
Db 74 KA-HOEGDVLVSVVAKSIISDVKIKGNSVITPEALKQNLGDANGFKVGDVLIREFKLENEFA 132
QY 142 TELTNQVISOYGYTETITVKTMLDGNRKLDMTFAEGKPARVVDINTIGNQHFSDADLI 201
Db 133 KSVKEHYASGVRYNATVEPVTNLPNNRAEILIOINEDDDRAKAKLASLTFKGNESVSSTLQ 192
QY 202 DVLAIKONKINPLSKADRYTOEKLVTSLLENRAKYLNAGFVRFETIKDAKLINEDKNRIF 261
Db 193 EQMELQPDSSWWKL-WGNKFEQAQPEKDLQSIIRDYLLNNGYAKAQITKTQVLDNEKTKVN 251
QY 262 VELSLHEGEQYRFGQTOFLGNLTYTQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLQTDLSRSARIIGNLGMSAELEPLLSALHLNDTFRSDIADVENAIKALGE 311
QY 321 DGYYYAQIRPVTRINDESRTVDVEYYIDPVHPVYVVRINFTGNEKTDQEVLRREMROLEG 380
Db 312 RGYGSATVNSVPDPDDANKTLAITLVVDAGRRLLTVRQLRPEGNIVSADSLRQEMRQOEG 371
QY 381 ALASNQIKLSRAKRLMTGTFKKHVTVDTR--PVPNSPDQVDNVFVBEQPSGSGSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVKVERNTGSIINFGIGY 429
QY 439 SQSGGVTFQFDVSONNPMGTGKHVNASFSSRSETREVYSLGTMNPYFTVNGVSQSLSGYR 498
Db 430 GTESGISYQASVQODNFGTGAAVSTAGTNDYGTSVNLGYTEPYFTKQGV--SLGNVVF 487
QY 499 KTRYDNK--NISNYLDSYGGSLSYGYPIDENQIRISFGL-----NADNTKLHGGRFMGI 550
Db 488 FENYDNKSDTSSNYKRTYGSNTLVGFPVNNENSYVGLGTYNKISNFALEYNRNLYI 547
QY 551 SNVKQLMADGGKIQVDNNGIPDPKHDYTYNAILGNWYSSLDREVFPFQGMHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYFPTKGVKASLGRVT 592
QY 609 VGFQDKTHQVYQGYNIYRPF-----IKKSVLRGVAKLGYGN-NLPFYENFVAGGYGS 660
Db 593 IPGSDNKKYKLSADVQGFYPLDRDLHLWVVSASAGYAN-GFGNKRLPFFQTYTAGGIGS 651
QY 661 VRGYDQSSLSGRSQAYLTARRQOTTLG-----EVVGGNALATFGSELILPLPFFKGDW 713
Db 652 LRGFAYGSIGPNA---IYAEYGNSGTGTPEKISSDVIGGNATATASAEILVPTPFVSDK 708
QY 714 I-DQVRPVIFIEGGQVFDTT-GMDKQTDLTQFDQPOATAEQNAKAANRPLLTDQKQLRY 771
Db 709 SQNTVTSIFVDAASVWMTKWSKDGLESVDLK-----RLPDYGKSSRIRA 755
QY 772 SAGVGATWYTPIGPLISISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVDEQFQFSIGGSF 797

RESULT 5
Y237_BUCAI STANDARD; PRT; 617 AA.
AC P57331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein BU237 precursor.
GN BU237.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbolic bacterium).
```


QY 676 YLTARRGOQTTLGVEVVGNGALATFGESEILPLPFKGDWIDQVRPVFIIEGGQVFTTGM 735
 Db 488 YAN-----GDLKAGSKLIT--GSLEYQVNVTKG-----WGAVFVDSGEA 525
 QY 736 KOTIDLTOKDQATAEQNAKAAANRPLLTQDKQLRYSGAGVATWTPICPLSLISYAKPLN 795
 Db 526 VSDIRSDPKT-----GTVGVWRWESPVPKIDFAVPA 560
 QY 796 KQNDQDTDTVOFQIG 810
 Db 561 DKDE--HGLQFYIG 572
 RESULT 8
 CG51_HUMAN
 ID CG51_HUMAN STANDARD; PRT; 469 AA.
 AC Q9Y512; OSNW85; O9UC09;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DI 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Protein CGI-51.
 GN CGI51.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
 RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
 RA Baguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
 RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
 RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
 RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
 RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
 RA Dodsworth S.J., Durrin R.M., Ellington A., Evans K.L., Fey J.M.,
 RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
 RA Graham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
 RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
 RA Kimberley A., King A., Laird G.K., Langford C.F., Leverisha M.A.,
 RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
 RA Matthews L., Mccann O.T., Mcclay J., McLaren S., McMurray A.,
 RA Milne S.A., Mortimore B.J., Odeh C.N., Pavitt R., Pearce A.V.,
 RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
 RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
 RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
 RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
 RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
 RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
 RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
 RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
 RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
 RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
 RA Lai H., Lao H.I., Lewis S., Lewis S., Lin S.P., Loh P., Malaj E.,
 RA Nguyen T., Pan H., Phan S., Qian Y., Ray L., Ren Q., Shaull S.,
 RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
 RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
 RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
 RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
 RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
 RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K.,
 RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
 RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
 RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
 RA Edlmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
 RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
 RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
 RA Khan A.S., Lane L., Tilahun Y., Wright H.;
 RT "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 RN [2]
 RP SEQUENCE FROM N.A.

RX MEDLINE=20272150; PubMed=10810093;
 RT Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
 RA "Identification of novel human genes evolutionarily conserved in
 RT Caenorhabditis elegans by comparative proteomics.";
 RL Genome Res. 10:703-713(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
 RA Ninomiya K., Iwavanagi T.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE UPF0140 (CGI-51) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL035398; CAB51401.1; -;
 DR EMBL; AK001087; AAD91498.1; -;
 DR EMBL; AF151809; BAA34046.1; -;
 FT CONFLICT 110 110 D -> G (IN REF. 3).
 FT CONFLICT 345 345 I -> V (IN REF. 3).
 FT CONFLICT 368 372 WAGGL -> LGRRW (IN REF. 2).
 SQ SEQUENCE 469 AA; 51976 MW; 4F686FB7ACB08DEF CRC64;
 Query Match 4.7%; Score 197.5; DB 1; Length 469;
 Best Local Similarity 22.0%; Pred. No. 0.00014;
 Matches 114; Conservative 82; Mismatches 212; Indels 111; Gaps 22;
 QY 334 INDESRVDVE---YYIDPVHPVYVRRINFTGNFKTQDEVLRRMQL---EGALASNQ 386
 Db 22 LGEEAEFVEVEPEAKOEILENKDVVQVHVFDGLGRTKDDIIICEIGDFKAKNLIEVVR 81
 QY 387 KTQLSRARLMRTGFEKHTV---DTRPVPS--PDQVDNVFVVE--OPSGSSTIAAGYSOS 441
 Db 82 KSHAREKLLRIGIFQVDVLDTCOGDALLPGLDVTFTVETLRILTSYNTWVGNNEG 141
 QY 442 GGVTFQFQVSNQNFMTGTHKVNASFSTRETVYSLGMTNP-----YFTVNGVSQSLS 494
 Db 142 SMV---LGLKLPNLLGRAEKVTFQFSYGTSETSYGLSFEKPRPGNFRNFSVN--LYKVT 196
 QY 495 GYV---RKTVDNKNISNYVLDVSGSLSYGPIDENQRIISFGLNADNTKLHGFRWG 549
 Db 197 GQFPSSURET-----DRGMAEYSFPIWKTS-----HTVKWEG 230
 QY 550 ISNVKQLMADGGKIQVDNNGIPDFKHDYTYTNAIILGNWYSSLDPRVFTQGMSSHVDLT 609
 Db 231 VNRRLGCLSRSTASFVAKESGHSLSKSLSHAMVI-----DSRNSSTLPRGALLKYNQEL 285
 QY 610 -GF--GDKTHQKVVOGNYIRPFIKKSVLRGAKLYGNNLP-----FYENFYAGGYG 659
 Db 286 AGYTGQDVSFIEDFELQNLKILDFSVFS--ASFWGMGLVPITGDKPSSSTADRYLGGPT 343
 QY 660 SVRGVYDQSSLSGPRSOAYLTARRQQOTTGLGVGGNATATFGSELILPLPEK---GDWIDQ 716
 Db 344 SIRGFSMSISGQSE-----GDYLGGEAYWAGGLHLYTLPFRPGQGGFGEL 390
 QY 717 VRPVIFIEGGQVFDATGMDKQTDIDLTFQDKDQATQNAKAAANRPLLTQDKQLRYSGAGV 776
 Db 391 FTTFELNAGNL-----CNLNYGSGPKAHIRKLAEC-----IRWSYGAG 429
 QY 777 ATW-YTPIGPLSIAKPLNKKNDQ--TDTVQFQIGSVF 813

Best Local Similarity 19.0%; Pred. No. 0.19;
Matches 199; Conservative 125; Mismatches 319; Indels 405; Gaps 55;

```
QY 9 FQVSAMTMMVMVMTSTHAQADFMADITITGLQVRITSLQSVLPFRLGQVSENQL-- 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 25 FGAASVLLGSLVLGAGQV--VKADETVAS--SEPIAS--SVAPASTEAEEAEKTN 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 ADGVKALYATGNFSDVQVYHQEGRIIYQVT-ERPLIAEINFEGNRLIPK---EGLQELK 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 AENTSAVATTS--TEVE---KAKAVLEQVTSSEPLLGL---GOKELAKTEDATLAKAIE 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 NA--GLAVGQPL---KQATQMIETELT-----NQYISOGYNTETI----- 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 131 DAQTKLAARAILADSEATVEQVEAQAAYKVAANEALGNELQKYTVDDLTAALDVIAPD 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 TVKQTM--DGNRVKLDMT-----FAEGKPARVVDINIIGNOHFSADLDLVLAIKDNK 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 191 TTASTLAVGDGEGTLLDSTTATPSMAEPNGAAIAP-----HILRTQDG- 234
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 211 INPLSKADRYTOEK--LVTSLENIJRAK--YLNA-----GFVRFEIKD-----AKLNINED 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 IKATSEPNWYTFESDYLSYKNKNMASSTYKGAEDVDAVIRYSLONDSSTTAVLAEL-VSRT 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 257 KNRFVETSLHEGROYREGOTQL---GNLTYT-----QAELEALLKFAEGFSGQAM 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 TGDVLEKTYITPEGESVTFSSHPTKVNANNSNITVYDTSASANTPGALKFSDNDVYSTI 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 307 L-----EQTNNISTKFGD-----DGYYY----- 325
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 IVPAYQINTTRYTESGKVLATYGLQIAGOVTPSSVRVFTGYDYVATTTKAVQGPYPK 413
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 326 -----AQIRPVTRINDESTVDVEYIIDPVHPVYVRINFTGNFTQDVELR 372
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 414 GTVYLAGTVQKTVQYKVIREIVENDQAVLKFYLD---PTKGEVDWRGT----- 461
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 373 REMROLEGALASNOKIQLSARLMRTGFFKHTVDTRPVNSPDQVDVNFVVEEQPSGS 432
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 462 -----DTGFIELT-----TSPYTKVGTIYD----- 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 TIAAGYSQSGGVTFQFDV-----SONNFMGTGKHVNASFSETR-----EYV- 475
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 -----YNINSKITAPFTIDPTKNVMVFKESEQESKYRVIQAWSGDETTKGYGKIYI 539
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 476 -----SLGMTNPYTVNGVQSLSG-----YYRK 499
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 540 ATQVWTKTLGTEWGFYSDQAGTKFNKNFPAQVQNTLRNATPATAVETTYIKES 599
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 500 TKY-----DNKNISNVLDYSYGSLSYGYPIDENORISFGLNADNT----- 540
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 600 SKYGDVIVEYDYGKQIVNSVVDTPKSLGTEYNTDVRPASLVAAADGTVYFYKEVKS 659
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 541 ---KLHGGRPMGINSVKQLMADGGKIQ---VDNNG-----IPDFKHDTYTYNAILGWNY 588
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 660 DSAKTTGTGVVAGTTTVKYVYKAGSVNVEVDINGKVIKAPVSDEK-----DAKPGYNI 713
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 589 SSLDRPVFPQOGSHSVDLTVGFGDKTHQKVYVQGNLYRFFIKSVLRG---YAKLGYGN 645
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 714 DT-----DL-----DOKLASITFEGKEY-----KLVPAGDYPVGVKGN 748
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 646 NLPFFYENFYAG-----YGSVRG-----YDQS-----SL 669
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 749 NLLEVGNNTAKGIDPTTGKTEAGVNKEVYVYRAVTVGSVVVYVKDTEGNVIKDPETDVS 808
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 670 GPRSQAYLT-----ARRGOOTTL-----GEVVGNALATGSELILPLPFK 710
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 809 APVGDAYTTTDDKKPNEIITKDSRYVLPFSKTDGEENGKVIETITVTVYVQKV----- 862
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 711 GDMIDQVRPVIFTEGGV---FDTTGMKQOTIDLTFQDKPQATAEQNAKANEPLLTQDK 767
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 863 ANWPEIPNVNPTDRPKVPYFPDPTED-EPIDPT-----TPGTNGEVPNIP----- 908
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 768 QLRYSAGVGATWTPYICPLSISYAKPLN 795
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 909 ---YVPG-----YTPVDPKDNTPKPID 928

RESULT 12
P3P_LACLC
ID P3P_LACLC STANDARD; PRT; 1902 AA.
AC P15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PIII-type proteinase precursor (EC 3.4.21.96) (Lactocepin) (Cell wall-
DE associated serine proteinase).
GN PTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RP STRAIN=SK11;
RC MEDLINE=89340435; PubMed=2760036;
RX Vos P., Simons G., Siezen R.J., de Vos W.M.;
RA "Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase.";
RL J. Biol. Chem. 264:13579-13585(1989).
CC -|- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -|- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some substrate preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J04962; AAA03533.1; ALT_SEQ.
CC PIR: A32634; A32634.
CC HSSP: P00782; 2SBT.
CC MEROPS: S08.019; -.
CC InterPro: IPR001899; Gram_pos_anchor.
CC InterPro: IPR003137; PA.
CC InterPro: IPR000209; Peptidase_S8.
CC Pfam: PF00746; Gram_pos_anchor; 1.
CC Pfam: PF02225; PA; 1.
CC Pfam: PF00082; Peptidase_S8; 3.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILASE_ASP; 1.
CC PROSITE: PS00137; SUBTILASE_HIS; 1.
CC PROSITE: PS00138; SUBTILASE_SER; 1.
CC PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
CC Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902 PIII-TYPE PROTEINASE.
FT DOMAIN 188 1876 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1877 1895 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 1896 1902 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 217 217 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
```

DT	01-OCT-1996 (Rel. 34, Created)
DT	01-OCT-1996 (Rel. 34, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	HYPOTHETICAL 51.8 kDa protein C17C9.06 in chromosome I.
DE	SPAC17C9.06.
OS	Schizosaccharomyces pombe (Fission yeast).
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;
OC	Schizosaccharomycetes.
OX	NCBI_TaxID=4896;
EN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=972;
RA	Murphy L., McDougall R., Jones L., Simpson I., McNeil A., Harris D.,
RA	Barrell B.G., Rajandream M.A., Walsh S.V.,
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC	-!- SIMILARITY: BELONGS TO THE UPF0140 (CGI-51) FAMILY.
CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
CC	EMBL; Z73099; CAA97352.1; -.
DR	Hypothetical protein.
KW	SEQUENCE 475 AA; 51762 MW; 8FEC5B5FF4B5DE15 CRC64;
SQ	-----
	Query Match 3.5%; Score 148; DB 1; Length 475;
	Best Local Similarity 19.1%; Pred. No. 0.12;
	Matches 102; Conservative 96; Mismatches 209; Indels 126; Gaps 24;
Qy	315 STKFGDDGYVYQIRPVTRINDESRTVDVEYY-----IDPVHPVYVRRINFTGNFKT 366
Db	: : : : : : : : : : : : : : : :
	7 STSPSPD-----IPAVNEESKLSAETFKSLSEILAENSTLPVGISSIRVTGAHHT 57
Qy	367 QDEVLRREMR-----QLEGALASNQIKLSRLMRGTGFKHVTVDTRVPVNS--- 414
Db	: : : : : : : : : : : : : : : :
	58 RPSIRVKLTCTLDTSKPAKRSLSLETLNAIQETGLMAFNVEYETANIKIDRASSVSG 117
Qy	415 PDQVDVNFVBEQP-----SGSSTIAAGYSQSGV--TFQFDVSNQNMFGTGRHVASFS 467
Db	: : : : : : : : : : : : : : : :
	118 DDDLDTVIQKEKPLRYVEGTDV-----GNVEGNVHANVLARNVFGAELLSGNS 169
Qy	468 -RSETREYISLGMPNPTFTVNGVSQSLSGYRKTKYDNKNISNVLDYSGGSLSYGYPID 526
Db	: : : : : : : : : : : : : : : :
	170 YGTRNRSTMSVNPETPVNADPKTRLRENGH--SNLRDNKSISSHDLLTKGITLSL----- 222
Qy	527 ENQIRISGLNADNTKLHGGRFPMGINSNVKQLMADGGKIQVDNGLIPDPKHDTYTYNAILGW 586
Db	: : : : : : : : : : : : : : : :
	223 QHODLWSGEHLISNL---LWRQVTHITETAYASPRVLEAGDSLKQSLSYTYTR----- 272
Qy	587 NYSSLDLRVPVPTQG--MHSVSDLT-VGF--GDKTHQVVYQGNIRYRPIKSV----- 634
Db	: : : : : : : : : : : : : : : :
	273 --DTRDLHLPPTKGDYVYRQILELAGFGFLGDSAFKSEFWG-----QKVALNLSRS 323
Qy	635 --LRGYAKLGYGNLL-----PFYENFYAGVSGYRGVYDQSSLGPRSQAYLTARRGQOTTL 687
Db	: : : : : : : : : : : : : : : :
	324 VSLSLSARIGALHSLNKKQVSLCDRFLMGGSTSLRGFSEDRIGPKD-----GRDS-- 373
Qy	688 GEVVGGNALATFGSELILPLPFGGDWTDQVRPIFIEGGQVDTTGMKMDKIDLTQFKDP 747
Db	: : : : : : : : : : : : : : : :
	374 ---LGGTAYMAFMSLLFLPLP-KVDASKPFRQLQFANAGLSNLTS-----PNP 418
Qy	748 QATAEQNAKANRPLLTDQKQLRYSACVGATWTPPTGPLSISYAKPLNKKQND 800
Db	: : : : : : : : : : : : : : : :
	419 CGTYK---SILSKPCI-----STGLGLVYATPAAARFELNFTLPIATTEKD 460

RESULT	13
YDF6_SCHPO	
ID	YDF6_1
AC	Q10471

ID	YEHB_ECOLI	STANDARD;	PRT;	826 AA.
AC	P33341;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical outer membrane usher protein yehB precursor.			
GN	YEHB_OR B2109.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / BHB2600;			
RA	Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,			
RA	Church G.M.;			
RL	Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12 / MGL655;			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.";			
RT	Sequence 277:1433-1474 (1997).			
RN	[3]			
RP	SEQUENCE OF 307-826 FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=97251358; PubMed=9097040;			
RA	Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,			
RA	Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,			
RA	Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,			
RA	Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,			
RA	Sivasubraman S., Tagami H., Takeda J., Takemoto K., Wada C.,			
RA	Yamanoto Y., Horiuchi T.;			
RT	"A 460-kb DNA sequence of the Escherichia coli K-12 genome			
RT	corresponding to the 40.1-50.0 min region on the linkage map.";			
RL	DNA Res. 3:379-392(1996)			
CC	-!- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF A FIMBRIAL			
CC	SUBUNIT ACROSS THE OUTER MEMBRANE (BY SIMILARITY).			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane			
CC	(By similarity).			
CC	-!- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U00007; AAA60473.1;			
DR	EMBL; AE000300; AAC75170.1;			
DR	EMBL; D90848; BAAL5975.1;			
DR	Ecogene; EGI1988; YehB.			
DR	InterPro; IPR000015; Fimb_usher.			
DR	Pfam; PF00577; Usher; 1.			
DR	PROSITE; PS01151; FIMBRIAL_USHER; 1.			
KW	Hypothetical protein; Outer membrane; Transmembrane; Fimbria;			
KW	Transport; Signal; Complete proteome.			
FT	SIGNAL	1	22	POTENTIAL.
FT	CHAIN	23	826	POTENTIAL.
FT				YEHB.
FT	DISULFID	809	825	POTENTIAL.
FT	SEQUENCE	826 AA;	92282 MW;	1FA541B1A21675CB CRC64;
CSQ				

Query Match 3.5%; Score 146; DB 1; Length 826;

Best Local Similarity 20.5%; Pred. No. 0.33;

Db 2 KRFLPAL-LSALMIAEVHAB--FTVSDIRVNGLQVSAGSFAALPLNVGETIDDOAL 58
Qy 67 ADGVKALYATNFSQVQVYHOEGRIIYQVTERPLIAEINFEGNRLIPKEGLQELKNAGL 126
Db 59 VOATRSFLKTFGFFQIQIDRGVNLVVTVERPSISSTIEIEGNAIKSIEDLLKGLKQSG 118
Qy 127 AVGOPLKOATQMETELTNOYISOGYNTETITVKOTMLDGNRVKLDMTFAEGKPARVVD 186
Db 119 AGEIFQATLEGVRNELQRYVAQGRYSABINAEVIPPQNRVALKININEGTVAAISH 178
Qy 187 INIIGNOHFSADLDIDVLAIK-DNKINPLSKADRYTOEKLVTSLLENRAKYLNAGFVFE 245
Db 179 INVGNIVFSEEDLTDLFELKTTNWSFFKNDKXAREKLSGDLERLSYLDRGYNMD 238
Qy 246 IKDAKLINEDKNRIFVEISLHEGEYRFGQTQFLGNTYTYQAELEALLKFKABEGFSQA 305
Db 239 IASTQVSIITPKKVIYITVINEGEKYITRDVKLTGDLKVPBEEVKRLLLVQKGQVESRK 298
Qy 306 MLEOTTNNISTKFGDDGYVYVIAQIRPVTRINDESRVDVEYIDVPVHYVRINFTNFK 365
Db 299 VMTTSLITRLNGEYITFANVNGVPEAHDDDKTVSVTFVDPGKRAYVNRINFRGNTK 358
Qy 366 TQDEVLRREMRLQEGALASNOIKIOLSRARLMRTGFFKHVTVTRPVNPSQDVDFVFE 425
Db 359 TEDEVLRREMRLQEGWASTYLDIOSKARLERLGYFKEVNVETPAVGTDDQVDVNSVE 418
Qy 426 EOPSGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHYNASFSRSETREVSGLMNTPYFT 485
Db 419 EOPSGSITASVGFASAGLIGGSISONNFLTGTGNKVSIGLTRSEYQTRYNFGVDPYT 478
Qy 486 VNGYSQSLSGYRKTGYD--NKNISNVLDYSGSLSGYGPIDENQRISEGLNADNTKLH 543
Db 479 VDGSLYNAFYKTDIDELVDVASTVNSLGAWSIGYPISTSLTYGLSVQRDQID 538
Qy 544 GGRPMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTYNAILGNWYSSLDPRVFPPTQGM 603
Db 539 TGRYT-VDEIYDFLDKEG---DN-----FTNFKSIGWSESTLNGKVLATRGHSQ 584
Qy 604 SVDL--TVGFGDKTHOKVYVYQGNIRYRFFIKKSVLRGAKLYGN-----NLFFYENFYA 655
Db 585 SLTETLTPGSDLSFYKIDYRGVFAPLDNYTMRFTHELGYGDGYGSTERLPFYENYA 644
Qy 656 GGXGSRVGYDQSSILGPRSQAYLTARRGOQTTL-----CEVYGGGNALAT 698
Db 645 GGNSVRGFKDSTLGPSTFSV-ARNPDGTPMKNOGPDCKRYTDPDQDPEAFEGGNILIT 703
Qy 699 FGSELILPLPKGDWIDQVPRVPIEGGQVDFDTGMDKQIDITQFKDPOATAEQNAKAA 758
Db 704 GGAELLPLPFVKDQ-RQLRTVLFWDVGSFTDCTKTNTNCDGIK-----749
Qy 759 NRPLLTQDKOLRYASAGVATWYPIGPLSISYAKPLNKKONDQTDVQFGISGVF 813
Db 750 -----TDNLASSVGVLTTWITLALGPLSFLATPIKKPDNAETQVQFSLGQTF 797

RESULT 2

Q9S341 ID Q9S341 PRELIMINARY; PRT; 797 AA.
AC Q9S341;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE ANTIGEN.
GN OMA.
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM.
RA Chatonnet-Marton P.I., Givaudan A., Ianois A., Boemare N.E.;
RT "Photorhabdus luminescens genomic region homologous to 4.0 minute

RT Escherichia coli region promotes pleiotropic phenotypes.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ236920; CAP51929.1; -
DR InterPro: IPR000184; Bac_surfAg_D15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
SQ SEQUENCE 797 AA; 88713 MW; 39E72E7ABD9C14F3 CRC64;

Query Match 29.8%; Score 1254; DB 2; Length 797;
Best Local Similarity 34.68; Pred No. 7e-61;
Matches 282; Conservative 161; Mismatches 334; Indels 38; Gaps 12;

Qy 14 MTMAVMVMVSTHAQAAD-FMANDITITGLQRTVIESQSVLFFRLQGVVSNQLADGVKA 72
Db 6 LLIASLLFGSAAAYGADGVVQDIFHEGLQRVAVGAALLNMPVRGDTVSDDIGRIHA 65
Qy 73 LYATGNSDVQVYHOEGRIIYQVTERPLIAEINFEGNRLIPKEGLQELKNAGLAVGQPL 132
Db 66 LFATGNFEDVRVLRDGNNTLIVQKERPTIASITSGNKSVDKMDLKQNLASHVRVGEAL 125
Qy 133 KOATVQMIETELTNOYISOGYNTETITVKOTMLDGNRVKLDMTFAEGKPARVVDINI 192
Db 126 DRTMLSNIERGLEDFYISVGKYNASVAVVTPLPNRVLDLKLVEAGVSAKIQIINIVGN 185
Qy 193 QHFSADLDIDVLAIKDNKINP---LSKADRYTOEKLVTSLLENRAKYLNAGFVFEIKDA 249
Db 186 KSFSSDELLANRFQLRDDV--PWNLTADQYKQKLGADLEALRSFYLDRGYARENDIST 243
Qy 250 KLAINEEDKNRIFVEISLHEGEYRFGQTQFLGNTYTYQAELEALLKFKABEGFSQAMLEQ 309
Db 244 QVSLTPDKKGIYVITNTEGDDQYKISGIDLNGNAGYQSEITKLAIEPGLSYNGTQVTK 303
Qy 310 TTNISTKFGDDGYVYVIAQIRPVTRINDESRVDVEYIDVPVHYVRINFTNFKTODE 369
Db 304 MENDIKNLGRYGVAYPRVMTQPEINDQDKTVLHVINDAGNRFYVKIRFSGNDTKDS 363
Qy 370 VLREMRQLEGALASNOIKIOLSRARLMRTGFFKHVTVTRPVNPSQDVDFVFEBOPS 429
Db 364 VLREMRQMERAWLGLSDLVELGKERLNLGYFETVDYETQRIPGSPQDVVYVYKVRNT 423
Qy 430 GSSTIAAGYSQSGVTFQFDVSONNFMGTGKHYNASFSRSETREVSGLMNTPYTVNGV 489
Db 424 GSLNFGVGTGTSVGFQIGAQDNLGTCNAGVINAASKNDYSTYAELESTDPFTINGV 483
Qy 490 QSLSGYSYRKTGYDNKNISNVLDYSGSLSGYGPIDENQRISEGLNADNTKLHGRFEMG 549
Db 484 SLGGRVYNDFRADDAELSGYTNQSYGSLGFLFPINENNSLNFGLNYIHNSL-SDMLPQ 542
Qy 550 ISNVKOLMADGGKIQVDNNGIPDFKHDTYTYNAILGNWYSSLDPRVFPPTQGMHSVD--L 607
Db 543 VAMVYLRSMGEKPDLESKA--EFKAD--DFALTMGWNTYNLDRGFFPTSGVKSTLNGKV 598
Qy 608 TVGFGDKTHOKVYVYQGNIRYRPF--IKKSVLRGAKLYGN-----NLFFYENFYAGYGS 660
Db 599 TIPGSDNEFYKVTLDTSAYYPINDRTWTWILGRSRLGYGDLGKGLKLPFTENYAGSSST 658
Qy 661 VRGYDOSSLGPRS-QAYLTARRGOQTTLGVEVVGNALATFGSELILPLPF-KGDWIDQVR 718
Db 659 VRGFRSNNGPKAIYLYKDGSPKSPRSRDAVGNAMAVASLELITPTPLPKDYSNSVR 718
Qy 719 PVTFIEGGQVDFDTGMDKQIDITQFKDPOATAEQNAKANRPLLTQDKOLRYASAGVAT 778
Db 719 TSFFIDSGTVMDTDWINDSAVM-----KSGIPDYSKPGNIRVSAGIALQ 762
Qy 779 WYTPIGPLSISYAKPLNKKONDQTDVQFGISGVF 813
Db 763 WMSPLGLPFVSYAKPIKDYEGDRSEQFQFNIGKTW 797

RESULT 3
Q9KPW0 ID Q9KPW0 PRELIMINARY; PRT; 803 AA.
AC Q9KPW0;

```
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SURFACE ANTIGEN.
GN VC252.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OC NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolaleva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDaniel L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL: AE004297; AAF95396.1; -.
DR TIGR: VC252.
DR InterPro: IPR000184; Bac_surfaG_d15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 803 AA; 89930 MW; 1998B3838F6C041D CRC64;

Query Match 28.6%; Score 1203; DB 16; Length 803;
Best Local Similarity 33.4%; Pred. No. 4.5e-58;
Matches 278; Conservative 155; Mismatches 328; Indels 72; Gaps 15;

QY 16 MAVVMVNSTHAQAD-FNANDITITGLQVYIESLQSVLPRLQGVVSENQADGVKALY 74
Db 8 LATLTVSVSANGAEKFFVQDIQGLQVVALGAALLKMPVRVGSVDSDVAIKALY 67
QY 75 ATGNFSDVQVYHQEGRITTYQTERPLIAEINFEGRNLPKCEQLQGLKAGLAVGQPLQ 134
Db 68 SSGNFEDVKVLRDGNTLAVQVKERTIASVSFSGNKALKEQLKONLEASSIRVGEALDR 127
QY 135 ATVQMIEFTLNQIYISQYNYETITVTKOTMLDGNRVKIDMTFAEGKPARVVDINIIGNOH 194
Db 128 TTLSNIEKGLDFYYSVGKYNATKAVVTPLPRNRADLKVFETEGSAKIQOINFGVQV 187
QY 195 FSDADLDVLAIKNKI--NPLSRADRYTQELVTSLENLAKYLNAQVFRFEIKDAKLIN 252
Db 188 FSDELLSRFNLVDVAMWNLAD-DKYKQVLADGIDIEALRTYLDRLGYLKFQVDSTQVA 246
QY 253 INEDKNRIFVEISLHEGQYRFGQTQFLGNLTYYQAELEALLKFAERGFSQAMLEQTTN 312
Db 247 ISPDKGVIYTLNLNEGEPYVSKVQFGLMGKAEFTSLIPFETIGYNGSAVTRLEE 306
QY 313 NISTKFGDDGYIAQIRPVTRINDESRTVDVEYIDPVHPVVRINTGFKQDEVLRL 372
Db 307 SVKVLGSEGYAPQVTRIPFEDDEKQVSLVHVHVEAGKRVVVRDREVGNNSTRDVLR 366
QY 373 REMRQLEGALASNOKIOLSRARLTGFKKHVTVDTRVPNSPDQVDVNFVVEEQSGSS 432
Db 367 REMRQMGESWLNSKDIETGKTRLRNLGFFETVEVQTVRVPSGSDVDLVYSKVEANSNV 426
QY 433 TIAAGYSQGGVTFQFQVSONNFMTGKHVNASFSSRSTREVSISLGMTNPYFTVNGVSQS 492
Db 427 NFGVGYGTESGVSFQVGLQODNFGLSGNRVGVNMINDYQKNLTLEYRDPYWNLDGVSIG 486
QY 493 LSGYRRTKYDNKNISNVLDVSGSLSYGYPIDENORISFGLNADNTKLRGPRMGISN 552
Db 487 GKVEYNQFEASEAGIVDYTNESYGTSLTWGFPFDELNRFEGIGYTHNKI--GNLTPYLQ 544
QY 553 VKQLMA-----DGGKIQVDNNGIPDFKHDYTYTNAILGNWYSSLDPRVPTQMSHS- 604
Db 545 VENFLAAQASNIDSGGNLLTDD-----FDINLSWTRNRLNNSYFFTAG-NHOR 591
```

```
QY 605 --VDLTVGCDKTHOKVYQGNIRPFIKKS-----VLRGYAKLGYGNN-----LP 648
Db 592 AFYKMTVPGSDAQYFKLQYDVRYQYPLTKKHEFTLLRG--RLGYNGYQOTDKGNLFP 649
QY 649 FYENFYAGYGVSRGYDQSSLSGPRS--QAYLTARRGQQTTLGVEVGNALATFGSELILP 706
Db 650 FYENFYAGGFTSLRGFGSNSAGPKAVYRDYSNGNSDSTATDDSVGNAIALASVELIVP 709
QY 707 LPFKGDWI--DOYRVPVIFIEGGQVFTTGMDKQTID-----LTFQKDPQATAEQNAANR 760
Db 710 TFFASEEARNQRTSIFYDMAVSMVMTDFYRGKADYGNQYVYDSDP-----756
QY 761 PLLTDQKQRLYSAGYGATWTPIGPLSISYAKPLNKKQNDQDTDTVQFOIGSVF 813
Db 757 -----TNYRSGYVALQWSPMGPLVFLSLAKPKIKYEGDDEFFFTTIGRTF 803

RESULT 4
Q9R2E3 PRELIMINARY; PRT; 785 AA.
AC Q9R2E3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 87.9 KDA PROTEIN.
GN YZZN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yanamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: Analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region."
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 2.4-4.1 min (110,917-193,643 bp) region."
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=82059454; PubMed=6272196;
RA An G., Bendjak L., Mametlak A., Friesen J.;
RT "Organization and nucleotide sequence of a new ribosomal operon in
RT Escherichia coli containing the genes for ribosomal protein S2 and
RT elongation factor Ts."
RL Nucleic Acids Res. 9:4163-4172(1981).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93077430; PubMed=1447125;
RA Yamanaka K., Ogura T., Niki H., Hiraga S.;
RT "Identification and characterization of the smba gene, a suppressor of
RT the mukB null mutant of Escherichia coli."
RL J. Bacteriol. 174:7517-7526(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Smallshaw J., Kelln R.;
RT "Cloning, nucleotide sequence and expression of the Escherichia coli
RT K-12 pyrH gene encoding UMP kinase."
RL Genetics 11:59-65(1992).
RN [6]
```

RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=94240115; PubMed=8183897;
 RA Janosi L., Shimizu I., Kaji A.;
 RT "Ribosome recycling factor (ribosome releasing factor) is essential
 RL for bacterial growth.";
 RN Proc. Natl. Acad. Sci. U.S.A. 91:4249-4253(1994).
 [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=91317739; PubMed=1860827;
 RA Shimizu I., Kaji A.;
 RT "Identification of the promoter region of the ribosome-releasing
 RL factor cistron (frr).";
 RN J. Bacteriol. 173:5181-5187(1991).
 [8]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90062117; PubMed=2684966;
 RA Ichikawa S., Kaji A.;
 RT "Molecular cloning and expression of ribosome releasing factor.";
 RL J. Biol. Chem. 264:20054-20059(1989).
 [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=86008268; PubMed=2995358;
 RA Icho T., Sparrow C.P., Raetz C.R.H.;
 RT "Molecular cloning and sequencing of the gene for CDP-diglyceride
 RL synthetase of Escherichia coli.";
 RN J. Biol. Chem. 260:12078-12083(1985).
 [10]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=91123198; PubMed=1991717;
 RA Hirvas L., Koski P., Vaara M.;
 RT "The ompH gene of Yersinia enterocolitica: cloning, sequencing,
 RL expression, and comparison with known enterobacterial ompH
 sequences.";
 RN J. Bacteriol. 173:1223-1229(1991).
 [11]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=91100302; PubMed=1987124;
 RA Dicker I., Seetharam S.;
 RT "Cloning and nucleotide sequence of the fira gene and the fira200(ts)
 RL allele from Escherichia coli.";
 RN J. Bacteriol. 173:334-344(1991).
 [12]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=88058790; PubMed=2824445;
 RA Crowell D., Reznikoff W., Raetz C.;
 RT "Nucleotide sequence of the Escherichia coli gene for lipid A
 RL disaccharide synthase.";
 RN J. Bacteriol. 169:5727-5734(1987).
 [13]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=88058791; PubMed=3316192;
 RA Tomaszewicz H.G., McHenry C.S.;
 RT "Sequence analysis of the Escherichia coli dnaE gene.";
 RL J. Bacteriol. 169:5735-5744(1987).
 [14]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=88139188; PubMed=3277952;
 RA Coleman J., Raetz C.;
 RT "First committed step of lipid A biosynthesis in Escherichia coli:
 RL sequence of the lpxA gene.";
 RN J. Bacteriol. 170:1268-1274(1988).
 [15]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;

RX MEDLINE=92380982; PubMed=1355089;
 RA Li S., Cronan J.;
 RT "The genes encoding the two carboxyltransferase subunits of
 RL Escherichia coli acetyl-CoA carboxylase.";
 RN J. Biol. Chem. 267:16841-16847(1992).
 [16]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90094229; PubMed=1688424;
 RA Zhou Z., Syvanen M.;
 RT "Identification and sequence of the drpA gene from Escherichia coli.";
 RL J. Bacteriol. 172:281-286(1990).
 [17]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90370122; PubMed=2203971;
 RA Ertani G., Delarue M., Poch O., Gangloff J., Moras D.;
 RT "Partition of tRNA synthetases into two classes based on mutually
 RL exclusive sets of sequence motifs.";
 RN Nature 347:203-206(1990).
 [18]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=93094132; PubMed=1459951;
 RA Gervais F.G., Drapeau G.;
 RT "Identification, cloning, and characterization of rcsF, a new
 RL regulator gene for exopolysaccharide synthesis that suppresses the
 RN division mutation fts284 in Escherichia coli K-12.";
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=94124004; PubMed=7904973;
 RA Allikmets R., Gerrard B., Court D., Dean M.;
 RT "Cloning and organization of the abc and mdl genes of Escherichia
 RL coli: relationship to eukaryotic multidrug resistance.";
 RN Gene 136:231-236(1993).
 [21]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=93011013; PubMed=1396599;
 RA Condon C., Phillips J., Fu Z., Squires C., Squires C.;
 RT "Comparison of the expression of the seven ribosomal RNA operons in
 RL Escherichia coli.";
 RN EMBO J. 11:4175-4185(1992).
 [22]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=79012454; PubMed=358189;
 RA Young R.A., Steitz J.A.;
 RT "Complementary sequences 1700 nucleotides apart form a ribonuclease
 RL III cleavage site in Escherichia coli ribosomal precursor RNA.";
 RN Proc. Natl. Acad. Sci. U.S.A. 75:3593-3597(1978).
 [24]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=93116053; PubMed=1474579;
 RA Cormack R., Mackie G.;
 RT "Structural requirements for the processing of Escherichia coli 5 S
 RL ribosomal RNA by RNase E in vitro.";

[illegible]

```
QY 432 STIAAGYSQGGVTFQFDVSONFMGTGKHVNASFSRSTREYVSLGTMTPYFTVNGVSQ 491
Db 432 FQVGLGYSKYGVTTVSQVLSQNNFLSGNRVSDASKRSYQDRYSFSYTNPFETDNGVSL 491
QY 492 SLGGYRKTKYDNKNISNYVLDYSGGSLSYGVYPIDENQRIISFGLNADNTKLHGGREMGIS 551
Db 492 GYNLAYOKLDYSDFNAAQYSKRMSGQTFIGPIPTENDTVSWVIGADSNQI--TTFPG-S 548
QY 552 NVQLMADGKQIQVDNNGIDPFKHDTTYNAIILGWNYSSLDLRVPFTQGHSHSDVLTGVF 611
Db 549 TPRAIID-----YIDAVG-----QRTFRWTELGWARDTRNDYFMPNLGMVQIGAEVTL 599
QY 612 GDKT--HOKVVYQGNIRPPIKKSIVLRGYAKLGVGN-----LPGY 650
Db 600 PGSTIKYKYNQISKVWPIIPALVNLTRLEVGVGDDYKSHTRILPDGTIVATASGLPFF 659
QY 651 ENFYAGYGSVGVYDSSLSQPRSOAYLTARRGOOTTLGEVVGNGALATFSGELILPLPFK 710
Db 660 ENFYAGTNSVRGFRDNTLGRSE--VTALYNQ---GOPLGGSEKTVGSTEMYFPKLF 713
QY 711 GDWIDQVRPVIIEGGGVFTTGMDKQTDIDLTQFKDPQATAEQNAKAANRPLLTQDKQLR 770
Db 714 S---PSARISAFDLDFGVNF-----GVNNFKA-----NELR 741
QY 771 YSAGVGATWTPICPLSISYAKPLNKKONQDQDTVQFQIGSVF 813
Db 742 ASSGVALLWRAPIGPISISYAFPIKKENDEIERLQTFEGQGF 784
RESULT 6
Q9JKH0 PRELIMINARY; PRT; 797 AA.
AC Q9JKH0
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMP85.
GN NMB0182.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
[1]
SEQUENCE FROM N.A.
RP STRAIN=MC58 / SEROGROUP B;
RC MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., Desoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Citterone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masiagnani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815(2000).
DR EMBL; AE002375; AAF40639.1; -.
DR TIGR; NMB0182; -.
DR InterPro; IPR000184; Bac_surfaG_D15.
DR Pfam; PF01103; Bac_surfaceA9; 1.
KW Complete proteome.
SQ SEQUENCE 797 AA; 88436 MW; B68BFC5A43D22EB8 CRC64;
```

```
Query Match 26.58; Score 1115; DB 16; Length 797;
Best Local Similarity 32.58; Pred. No. 3.le-53;
Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;
QY 10 QVSAMTAVMMVMTAAQADFAMNDITITGLQRTVIESLQSVLFFRLGQVVSQNLADG 69
Db 2 KLKQIASALMLGMSPLALADFTIQDIRVEGLQRTPESTVFNLPVKVGDYNDTHGSAI 61
```

```
QY 70 VKALYATGNFSDVOYVHOEGRIIYQVTERPLIAINEFGRNLIPEKGLQEGKKNAGLAVG 129
Db 62 IKSLYATGFFDDVRVETADGOLLTVIERPTIGSLNITGAKMLQNDALIKKLESFGLAQ 121
QY 130 OPLQKQATVQMLETELTNQYISOGYYNTEITVKQTMLDGNRVKLDMTAEKGPARGVDINI 189
Db 122 QYFNOATLNOAVAGLKEBYLGRGKLNIOITPKVKTLARNRVDDIITIDEGSAKITDEF 181
QY 190 IGNOHFSDADLIDVLAJDKNKI-NPLSKADRYTOEKLVTSLENIRAKYINAGFVRFEIKD 248
Db 182 EGNQVYSDRKLMRQNSLTGEGITWLTFRSQFNEQKFAQDMKQVTDYFQNNGYDFRILD 241
QY 249 AKLNINEDKNRIFVEISLHEGEYRFQGTQFLGNLT-YTQAEALALKFAEFGFSQAML 307
Db 242 TDIQTNEKTKQTIKTIVHEGGRFGWKVSIEGTDNEVPKAELEKLLTMKPGKMYERQOM 301
QY 308 EOTTNNISTKFGDGGYVYQAIQIRVTRINDESRSTVDVEYIDVPVHVYVRRINFTGNFKTQ 367
Db 302 TAVLGEIONRMGSAGYAISEISVQPLPNAETKTVDVLIHEPGRKIYVNEIHTGNKTR 361
QY 368 DEVLRRERMOLEGLASLNQIKQLSRARLMRTGFFKHVTVTRPVPNSPDQDVNFVVEEQ 427
Db 362 DEVVRELQMESAPYDTSKLQRSKERVLLGYFDNVQFQDAVPLAGTDPKVDLNMSLTER 421
QY 428 PGSSTIAAGYSQGGVTFQFDVSONFMGTGKHVNASFSRSTREYVSLGTMTPYFTVN 487
Db 422 STGSLDLSAGVWQDTGLVMSAGVSDNLFGTGKSAALRASRKTTLNGSLSTFDYFTAD 481
QY 488 GVSQSLSGYR-KTKYDNKNISNVLDYSGGSLSYGVYPIDENQRIISFGLNADNTKLHGG 545
Db 482 GVSIGVDYVGRAPPRKASTSIKQYKTTAGAGIRMSVPVTEYDRVNEGLVAEHLTVN-- 539
QY 546 RFMGISNVKQLMADGGKIQVDNNGIPDPFKHDYTYTNALIGWNYSSLDLRVPFTQGHSHV 605
Db 540 --TYNKAPHYADFIKKYKTDG-TDGSFKGWLYKGVGWGRNKTDLSALWPTRGYLTGV 595
QY 606 DLTGVF-GDK--THQVYVQGNIRPPIKKSIVLR-----GYAKLGV--NNLPFFYENFVAG 656
Db 596 NAEIALFGSKLQYYSATHNQTWTFPLSKTFTLMLGGEVGIAG-GYGRTKETIPFFENFYGG 654
QY 657 GYGSVGVYDSSLSQPRSOAYLTARRGOOTTLGEVY--GGNALATFSGELILPLPFGDWI 714
Db 655 GLGSVGVYESGTILGPK-----VYDEYGEKISYGGNKKANYSAEILLPMPGAKD-A 703
QY 715 DQVRPVIIEGGQVFDTTGMDKQTDIDLTQFKDPQATAEQNAKAANRPLLTQDKQLRYAG 774
Db 704 RTVRLSLFADAGSVMDGKTYDDNSSAT-----GGRVQNIYCGAGNTHKSTFTNELRYSAG 758
QY 775 VGATWYTPIGLPLSISYAKPLNKKONQDQDTVQFQIGSVF 813
Db 759 GAVTWLSPLGPMKFSYAYPLKKKKPEDEIQRFQFQJGTTF 797
RESULT 7
Q9JX31 PRELIMINARY; PRT; 797 AA.
AC Q9JX31
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMP85.
GN OMP85 OR NMA0085.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
[1]
SEQUENCE FROM N.A.
RP STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RC MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Mungall K., Quail M.A.,
```


RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RT Whitehead S., Spratt B.G., Barrell B.G.;
"Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491";
RL Nature 404:502-506(2000).
DR EMBL; AL162752; CAB83401.1; -;
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 797 AA; 88404 MW; 65DE47E00C9E1DIF CRC64;

```
Query Match      26.4%; Score 1111; DB 16; Length 797;
Best Local Similarity 32.4%; Pred. No. 5.1e-53;
Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

Qy 10 QVSAMTAMVMVMSHTAAQADFMANDITITGLQRTVIESLQSVLPFRGLQGVVSENQADG 69
Db 2 KLKQIASALMWLGISPLALADFTIQDIRVEGLQRTPESTVFVNLVPVKVGDYNDTHGSAI 61

Qy 70 VKALYATGNFSDVQVYHQEGRIYQVTERPLIAEINFEGRNLIPKQEGLEKLNAGLAVG 129
Db 62 IKSLYATGFDDVRVETADGQLLTVERPTIGSLNITGAKMLQDAIKKNLESFGLAQS 121

Qy 130 QPLKQATVOMIETELTNQYISQYYNTEITVKQTMLDGNRVKLDMTFAEGKPARVVDINI 189
Db 122 QYFNQATLNQAVAGLKEEYLGKGLNIQITPKVTKLARNRVDITIDEGSKAKITDIEF 181

Qy 190 IGNOHFSADLDIVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRAKYLNAGFVFEIKD 248
Db 182 EGNQVYSDRKLQMSLREGGIWTLTRSNQFNEQFAQDMKVTDFYQNNGYDFRILD 241

Qy 249 AKLNINEDKNRIFVEISLHEGEQYRFGQTFGLNLT-YTQAEALALKFAEGFSQAML 307
Db 242 TDIQTNEDKTKOTIKITVHEGGRFWMGVKYSIEGDTNEVPKAELEKLLTKPKWYERQOM 301

Qy 308 EOTTNNISTKFGDDGYVYQAQIRPVTRINDESTVDVEYIDPVHPVYVRRINFTGNFKTQ 427
Db 362 DEVVRELRQMSAPYDTSKQSKERVLLGYFDNVQFQFVPLAGTDPKVDLNNMSLTER 421

Qy 428 PGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHYNASFSESTREVSILGMTNPFTVN 487
Db 482 GVSQSLSGYYR--KTKYDNKNISNVLDYSYGSLSYGYPIDENQRIISFGLNADNTKLHGG 545

Qy 546 RFMGINSNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFTQGMHSV 605
Db 540 ---TYNKAPKHVADFIKKYKTDG--TDGSFQKWLKYGTVGWRNKRTDSALMPTRGLYTCV 595

Qy 606 DLTGVF-GDK--THQKVYVQGNIIYRPFIKKSVLR-----GYAKLGYG--NNLPFFYENFYAG 656
Db 596 NAEIALPGSKLQYYSATHNQWTFPPLSKTFTLMLGGEVGIAG-GYGRTKKEIPFFENFYGG 654

Qy 657 GYGSVRGVDQSSLSGPRSQAYLTARRGQQTTLGEVV--GGNALATGSELILPLPKPGDWI 714
Db 655 GLGYSVRGESLTGPK-----VYDEXGEKISYGNKNKANYSAELLFPMPGAQD-A 703

Qy 715 DQVRVPIEGGVDFDTGMDKQIDLTQFKDPQATAEQNAKAANRPLLTQDKQLRYSG 774
Db 704 RTVRLSLFADAGSVWDGTYTDNNSSAT-----GGRVQNIYGAGNTHKSTFTTNELRYSG 758

Qy 775 VGATWTPIGPLSISYAKPLANKQNDQTDTFQFQIGSVF 813
Db 759 GAVTWLSPLGPMKFSYAYPLKKKKPDEIQRFQGLGTTFF 797
```

```
RESULT      8
O30912
AC O30912 PRELIMINARY; PRT; 797 AA.
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMP85.
GN OMP85.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HH.
RA Manning D.S., Reschke D.K., Judd R.C.;
RT "Omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis are similar to Haemophilus influenzae D-15-Ag and Pasteurella multocida Oma87";
RL Microb. Pathog. 23:0-0(1998).
DR EMBL; AF021245; AAC17599.1; -;
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
SQ SEQUENCE 797 AA; 88539 MW; CF911B5F70B999CF CRC64;

Query Match      26.4%; Score 1109; DB 2; Length 797;
Best Local Similarity 32.4%; Pred. No. 6.6e-53;
Matches 265; Conservative 154; Mismatches 362; Indels 38; Gaps 14;

Qy 10 QVSAMTAMVMVMSHTAAQADFMANDITITGLQRTVIESLQSVLPFRGLQGVVSENQADG 69
Db 2 KLKQIASALMWLGISPLAFADFTIQDIRVEGLQRTPESTVFVNLVPVKVGDYNDTHGSAI 61

Qy 70 VKALYATGNFSDVQVYHQEGRIYQVTERPLIAEINFEGRNLIPKQEGLEKLNAGLAVG 129
Db 62 IKSLYATGFDDVRVETADGQLLTVERPTIGSLNITGAKMLQDAIKKNLESFGLAQS 121

Qy 130 QPLKQATVOMIETELTNQYISQYYNTEITVKQTMLDGNRVKLDMTFAEGKPARVVDINI 189
Db 122 QYFNQATLNQAVAGLKEEYLGKGLNIQITPKVTKLARNRVDITIDEGSKAKITDIEF 181

Qy 190 IGNOHFSADLDIVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRAKYLNAGFVFEIKD 248
Db 182 EGNQVYSDRKLQMSLREGGIWTLTRSNQFNEQFAQDMKVTDFYQNNGYDFRILD 241

Qy 249 AKLNINEDKNRIFVEISLHEGEQYRFGQTFGLNLT-YTQAEALALKFAEGFSQAML 307
Db 242 TDIQTNEDKTKOTIKITVHEGGRFWMGVKYSIEGDTNEVPKAELEKLLTKPKWYERQOM 301

Qy 308 EOTTNNISTKFGDDGYVYQAQIRPVTRINDESTVDVEYIDPVHPVYVRRINFTGNFKTQ 367
Db 302 TAVLGEIQNRMGSAIYSEISVQPLPNAETKTVDFVLHIEPGRKIYVNEIHTGNKTR 361

Qy 368 DEVLRRERQLEALASNQIQLRSARLMRTGFFKHVTYDTRPVNPSDQVDVNFVEEQ 427
Db 362 DEWVRRELQMSAPYDTSKQSKERVLLGYFDNVQFQFVPLAGTDPKVDLNNMSLTER 421

Qy 428 PGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHYNASFSESTREVSILGMTNPFTVN 487
Db 422 STGSLDLGAGVQDFTGLVMSAGVSQDNLFCTGKSAALRASRSKTTLNGSLSEFTDPTAD 481

Qy 488 GVSQSLSGYYR--KTKYDNKNISNVLDYSYGSLSYGYPIDENQRIISFGLNADNTKLHGG 545
Db 482 GYSLGYDYVYKAFDPRKASTSIKQYKTTTAGAGIRMSVPVTEYDRVNEGLVAEHLTVN-- 539

Qy 546 RFMGINSNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFTQGMHSV 605
Db 540 ---TYNKAPKHVADFIKKYKTDG--TDGSFQKWLKYGTVGWRNKRTDSALMPTRGLYTCV 595

Qy 606 DLTGVF-GDK--THQKVYVQGNIIYRPFIKKSVLR-----GYAKLGYG--NNLPFFYENFYAG 656
Db 596 NAEIALPGSKLQYYSATHNQWTFPPLSKTFTLMLGGEVGIAG-GYGRTKKEIPFFENFYGG 654

Qy 657 GYGSVRGVDQSSLSGPRSQAYLTARRGQQTTLGEVV--GGNALATGSELILPLPKPGDWI 714
Db 655 GLGYSVRGESLTGPK-----VYDEXGEKISYGNKNKANYSAELLFPMPGAQD-A 703

Qy 715 DQVRVPIEGGVDFDTGMDKQIDLTQFKDPQATAEQNAKAANRPLLTQDKQLRYSG 774
Db 704 RTVRLSLFADAGSVWDGTYTDNNSSAT-----GGRVQNIYGAGNTHKSTFTTNELRYSG 758

Qy 775 VGATWTPIGPLSISYAKPLANKQNDQTDTFQFQIGSVF 813
Db 759 GAVTWLSPLGPMKFSYAYPLKKKKPDEIQRFQGLGTTFF 797
```


Db	255	K	I	N	E	S	O	N	I	S	E	M	R	I	G	D	T	Q	K	D	N	E	L	N	Q	L	T	H	F	K	A	Q	L	F	R	K	T	S	L	S	I	E	S	I	Q	I	K	O	I	L	G	D	R	G	Y	314
QY	324	Y	A	Q	I	R	P	V	T	R	I	N	D	E	S	T	V	E	Y	I	D	P	V	H	P	V	Y	R	R	I	N	F	T	Q	D	E	V	L	R	R	E	M	R	Q	E	G	A	L	A	383						
Db	315	G	S	A	K	V	D	L	I	P	K	F	E	E	D	H	T	V	I	N	F	I	D	A	G	R	I	Y	R	K	I	F	E	C	N	D	T	A	D	S	T	L	R	R	E	M	R	Q	E	G	A	W	L	374		
QY	384	S	N	O	K	I	O	L	S	R	A	R	L	M	R	T	G	F	F	K	H	V	T	D	T	R	P	V	N	S	P	D	Q	V	D	V	N	F	V	E	Q	E	S	G	S	T	I	A	A	G	Y	S	Q	S	G	443
Db	375	S	T	S	A	V	S	L	A	K	S	R	L	R	E	T	F	E	T	V	E	M	S	M	P	T	V	K	N	T	D	Q	D	V	I	I	K	I	K	E	R	N	T	G	S	I	N	F	G	Y	G	S	G	434		
QY	444	V	T	F	O	D	V	S	O	N	N	E	M	G	T	K	H	N	A	S	R	S	E	T	R	E	V	T	S	L	G	M	T	N	P	E	T	V	N	G	S	O	L	S	G	Y	R	K	T	Y	D	503				
Db	435	L	S	T	N	A	G	I	T	Q	D	N	F	L	G	M	S	S	L	G	L	G	S	R	N	D	T	S	T	N	V	N	L	S	T	E	P	T	F	K	D	G	V	--	S	L	G	N	I	F	E	D	Y	492		
QY	504	N	--	K	N	I	S	N	V	L	D	S	G	S	L	S	Y	G	P	I	D	E	N	O	R	I	S	F	L	G	N	A	D	T	K	L	H	G	R	F	M	G	I	S	N	V	K	O	I	L	M	A	D	G	560	
Db	493	N	S	A	K	A	S	A	A	Y	K	R	K	T	Y	G	A	S	G	T	G	L	G	P	P	V	D	E	N	S	Y	L	G	Y	T	H	D	K	L	--	R	N	V	E	R	E	T	R	E	K	Y	N	S	549		
QY	561	G	K	T	O	V	N	D	N	G	I	P	D	K	H	---	D	Y	T	T	N	A	I	L	G	W	N	S	S	L	D	R	P	V	P	T	Q	G	M	S	H	S	V	D	--	L	T	V	G	F	G	D	K	614		
Db	550	M	K	F	P	I	N	---	P	Q	N	S	H	Y	D	R	I	Q	S	A	D	F	D	L	S	F	G	W	N	N	L	R	G	F	T	A	G	S	S	A	N	I	S	G	L	T	L	P	G	S	D	605				
QY	615	T	H	O	K	V	V	O	G	N	I	R	P	F	I	K	--	K	S	V	L	R	G	Y	A	K	L	G	Y	---	G	N	N	L	P	F	E	N	P	F	A	G	Y	G	S	V	R	G	Y	D	Q	S	667			
Db	606	K	Y	T	O	V	G	T	N	F	S	G	Y	I	P	L	N	S	E	H	K	W	I	A	T	G	G	L	A	Y	N	S	F	G	G	K	E	V	P	F	Q	L	S	A	G	M	S	L	R	G	F	A	G	665		
QY	668	S	L	G	P	S	O	A	Y	L	T	A	R	G	O	T	T	L	G	E	V	G	G	N	A	L	T	F	E	S	L	I	P	L	P	F	K	G	D	--	W	I	D	Q	V	R	P	V	I	F	I	E	G	726		
Db	666	S																																																						

```

Query Match      25.0%; Score 1050; DB 2; Length 792;
Best Local Similarity 31.5%; Pred. No. 1.1e-49;
Matches 258; Conservative 159; Mismatches 334; Indels 68; Gaps 20;

Qy 23 STHQAADFNDADITITGLQRTVTIESQSVLPFRLGQVVSFENQLADGVKALYATGNFSDV 82
      :|||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

```

Db	14	TTTTFAAFFVAKDIRVDCVGQDLEQQIRASLPRAGORVTDNDVANIVRSLVFSGRFDY 73
Qy	83	QVYHQEGRI-IYQVTERPLIAEINFEGNRLIPKEGLQEGLNAGLAVGQPLKQATQMIE 141
Db	74	KA-HOEGDVLVSVVAKSIISDKVIKSIIPDTEALKNDLNDANGFKVGDILIREKLNFEA 132
Qy	142	TELTNQIYSQGYNYETIVTQMTLDGNRVKLTMTFAEGKPARVVDINIGNQHFSDADLI 201
Db	133	KSVKEHTVASGRYNATPEIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSTLQ 192
Qy	202	DVLAIKDNKINPLSKADRYTOEKLVTSLENLRKALYNAGFVRFEIKDAKLINEDKNRIF 261
Db	193	BOMELQDPDSWKWL-WGNKFCGAOFEKDQAIROYDYLNNGYAKAQTFTVDOLNDEKTKVN 251
Qy	262	VEISLHEGEQYRFQOTFLGNLTATQAEALALK-FKAEEGFSQAMLEQTTNNISTKFGD 320
Db	252	VTIDVNEGQYDLRSARIIGNLGGMSAELEPLLALHLNDFRRSDIADVENAIKAKLGE 311
Qy	321	DGYIYAQIRPVTRINDSRTPVDEYYIDPVHPVYVRINFTGNPKTQDEVLRREMQLGEG 380
Db	312	RGYGNNTVNSVPDFDDANKTLAITFVVDAGRRLLTVHOLRFEQNTVSADSTLROEMRQEG 371
Qy	381	ALASNOKIQLSRARLMTGTFKKHVTVDTR--PVPNSPDQVDVNFVVEEQSGSSTIAAGY 438
Db	372	TWYNSQLVEGKIRLDRKTGFEE--TVENRIDPINGSNDEVDVYVYKERNWTGSIKFGY 429
Qy	439	SQSGGVTFQFDVSONNFMGTGKHKNASFSRSETREYVSLGNTNPFYTVNGVSQSLSGYR 498
Db	430	GTESGISYQASVKODNLFELGTCAAVSIAGTKNDYCTSVNLGYTEPYTKDGV--SLGNNVF 487
Qy	499	KTKYDNK---NISNVLDSTYGGSLYSYGPIDENQRIISFGL----NADNTKLGGRFMGI 550
Db	488	FENTDNSKSDTSNNYKRTTYGSNNVTLGFPYNNENNSYVYGLGHTYNTKISNPALEYNRNLYI 547
Qy	551	SNVKOLMADGKIQVDNNGIPDFKHDYTTYNAILGNWYSSLDRPVPPTQGMHSVD--LT 608
Db	548	QSMK-----FKNGIKTNDDFS-----FGWNYSNLNRGYPTKGVKASLGRVTV 592
Qy	609	VGFQDKTHQKVYOGNIYRPF-----IKKSILRGYAKLGYGN-NLPFYENFYAGYGS 660
Db	593	IPGSDNKKYKLSADVQGFPLDRDHLVWSAKASAGYAN-GFGNKRLLPFYQTYTAGSIG 651
Qy	661	VRGVDDQSSLGRPSQAYLTARQGQTTL----GEVVGNALATFGSELILPLPKGDMW-D 715
Db	652	LRGEAYGISGPN-----IYQGNKNKFISSDVIIGNATIASAELIIVTPFVSDKSON 706
Qy	716	QYRVPITFEGQVFDTT-GMDKQITDITQPKDQATAEQNAKANRPLLTDQDKLRYASAG 774
Db	707	TVRTSLFVDAASVNTKKWSDKNGLSENVLKD-----LPDYKSSRIRASTG 753
Qy	775	VGATWYTPIGPLSISYAKPLNKKNDQTDTVQFOIGSVF 813
Db	754	VGFQWSPIGPLVFSYAKPIKKYENDDDVEQFQFSIGSF 792
RESULT	12	
Qy	ID	PRELIMINARY; PRT; 791 AA.
Qy	AC	Q9CJLI;
Qy	DT	01-JUN-2001 (TrEMBLrel. 17, Created)
Qy	DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
Qy	DE	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
Qy	DE	HYPOTHETICAL PROTEIN PM1992.
Qy	GN	PM1992
Qy	OS	Pasteurella multocida.
Qy	OS	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Qy	OC	Pasteurella.
Qy	NCBI_taxid	747;
Qy	RN	[1]
Qy	RP	SEQUENCE FROM N.A.
Qy	RC	STRAIN-PW70;
Qy	RA	MEDLINE=21145866; PubMed=11248100;
Qy	RX	May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).
DR EMBL; AF06235; AAK04076.1; -;
DR InterPro; IPR001184; Bac_surfaG_D15.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF01103; Bac_surfaceAg; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein; complete proteome.
SQ SEQUENCE 791 AA; 87761 MW; 024550DD8F99CD82 CRC64;

```
Query Match      25.0%; Score 1049; DB 16; Length 791;  
Best Local Similarity 31.3%; Pred. No. 1.3e-49;  
Matches 258; Conservative 162; Mismatches 344; Indels 60; Gaps 17;  
  
QY 14 MTAVMMVMSTHAAQADFMANDITITGLQRTVTESLQSVLPFRLGQVVSENQADGVKAL 73  
DB 4 LLTASLLFGSTTAFAPFVVKDIRVQGVQAGTEGSLVATLPVRVGQRATDNDIANVVRKL 63  
  
QY 74 YATGNFSDVOVYHQEGRIIYQVTERPLIAEFNFGNRLIPKEGLOEGLKNAGLAVGQPLK 133  
DB 64 FLSSQYDDVKASREGNTLVVTMPKPVISNVWIDGNKSIDPEAIKQNLDAANGFKVGDVNL 123  
  
QY 134 QATVQMETELTNQYISQGYNTETITVKQTMLDGNNRVKLDMTFAEGKPARVVDINIIGNQ 193  
DB 124 RAKLEEFKRGIVEHYNSVGRYNAKVEAIVNTLPNNSAEIKIQINEDDVALFKEITFEQNG 183  
  
QY 194 HFSADADLDVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYNAGFVRPEIKDAKINI 253  
DB 184 AFSSKLEQDQMETDAWKKLF-GNKFDTQTFNKDLETLSYLLDRGYAQFILDITDKL 242  
  
QY 254 NEDKNRIFVEISLHEGQYRFGQTFIGNLTYYQAELEALLKFAEKG-FSQAMLEQTTN 312  
DB 243 SDDKKEARVLIKVEKGLDYTVKSARILGVDGMSAELAPILATQLNGLFRRANVLEVEQ 302  
  
QY 313 NISTKFGDGGYVYQAIQRPVTRINDESRITVDVEYIIDPVHPVYVRRINFTGNFTQDEVLR 372  
DB 303 RIKSKLGERGYATAQNVNHTFDEQDKTISLDFIVEAGKSYTVRQIRFEGNTSSADSTLR 362  
  
QY 373 REMRQLEGALASNOKIQLSRARLMRTGFFKHTVTDTRPVPNSPDQVDVNVFVEEQSGSS 432  
DB 363 QEMRQEGAWLSSELVELGLRLDRDTGFESVETKTEAIPGS-DQVDVIYKVKERNITGSI 421  
  
QY 433 TIAAGYSQSGGVTFQPDVSONNFMGKHKVNASFSETSRETVSLGNTNPFYTVNGVSQS 492  
DB 422 NFGIGYGTESGLSYQASIKQDNFLGMGSSISLGGTRNDYGTITNLGYNPEYFTKDGV--S 479  
  
QY 493 LSGYIRKTKYDKNINIS--YVLDSYGGSLSYGYPIDENQRISEGLNADNTKLHGGRFMG 549  
DB 480 LGGNVFEEDSSKNTSAAYGRTSYGGNLTGLFPVNNENNSYILGVGYTYTNKL----- 532  
  
QY 550 ISNV-KOLMADGGKIQVDNNGIPDFK-HDYTTYNAILGMWYSSLDPRVPFTQGMSSHSDV- 606  
DB 533 -KNIAPEYNRDLRYQSMKYNDWSTFKSHD--FDLSFGWYNSLNRGYFPTKGVANIG 588  
  
QY 607 -LIVGSGDKTHQKVVQGNTRYRPF-----TKKSVLRGYAKLYGNNLPFTNFY 654  
DB 589 RVTIQSDNNKYIKLNABEAQGFYPLDREHGWLVSSRSISASFADGFG-----GKRLFFYQYIS 644  
  
QY 655 AGGYSGVRGYDQSSLPQRSQAYLTARRGQQTTL--GKVGGNALATFGSELILPLPKGD 712  
DB 645 AGGIGSLRGFAYGAIGNA-IYRTQCPDSYCLVSDSDVIGGNWAVTASTELIYVTFPVAD 703  
  
QY 713 W-IDQVRPVIIFIEGGQVFDFTGMDKQITIDLPKDPQATAEQNAKAA--NRPLLTQDOKL 769  
DB 704 KNQNSVRTSLFVDAASVWNT-----RWKAEDKAKFAKLVNPDYSDPSRV 747  
  
QY 770 RYAGVATHYTTIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813  
DB 748 RASAGVALQWQSPIGPLVFSYAKPLKKYQDGEIEQFQISGGTF 791
```

RESULT 13

Q51930
ID Q51930 PRELIMINARY; PRT; 789 AA.
AC Q51930; Q51921;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE ANTIGEN OMA87.
GN OMA87.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_taxid=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PBA100;
RA Ruffolo G.C., Adler B.;
RL Infect. Immun. 64:0-0(0).
RN [2]
RP SEQUENCE OF 750-789 FROM N.A.
RC STRAIN=9222;
RX MEDLINE=95369730; PubMed=7642134;
RA Delanarche C., Manoha F., Behar G., Houlgatte R., Hellman U.,
RA Wroblewski H.;
RT "Characterization of the Pasteurella multocida skp and fira genes.";
RL Gene 161:39-43(1995).
DR EMBL; U60439; AAC44600.1; -;
DR EMBL; X74357; CAA52399.1; -;
DR InterPro; IPR001184; Bac_surfaG_D15.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF01103; Bac_surfaceAg; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 789 AA; 87437 MW; DC1D004AAFB2CD9E CRC64;

```
Query Match      23.8%; Score 1000; DB 2; Length 789;  
Best Local Similarity 30.8%; Pred. No. 6.4e-47;  
Matches 254; Conservative 164; Mismatches 343; Indels 64; Gaps 21;  
  
QY 14 MTAVMMVMSTHAAQADFMANDITITGLQRTVTESLQSVLPFRLGQVVSENQADGVKAL 73  
DB 4 LLTASLLFGSTTAFAPFVVKDIRVQGVQAGTEGSLVATLPVRVGQRATDNDIANVVRKL 63  
  
QY 74 YATGNFSDVOVYHQEGRIIYQVTERPLIAEFNFGNRLIPKEGLOEGLKNAGLAVGQPLK 133  
DB 64 FLSSQYDDVKASREGNTLVVTMPKPVISNVWIDGNKSIDPEAIKQNLDAANGFKVGDVNL 123  
  
QY 134 QATVQMETELTNQYISQGYNTETITVKQTMLDGNNRVKLDMTFAEGKPARVVDINIIGNQ 193  
DB 124 RAKLEEFKRGIVEHYNSVGRYNAKVAIVNTLPNNSAEIKIQINEDDVALFKEITFEQNE 183  
  
QY 194 HFSADADLDVLAIKDNKINPLSKADRYTQEKLVTSLENRAKYNAGFVRPEIKDAKINI 253  
DB 184 AFSSKLEQDQMETDAWKKLF-GNKFDTQTFNKDLETLSYLLDRGYAQFILDITDKL 242  
  
QY 254 NEDKNRIFVEISLHEGQYRFGQTFIGNLTYYQAELEALLKFAEKG-FSQAMLEQTTN 311  
DB 243 SDDKKEPCL-IS-EGDLTYV-KTRVSGGMWGMGMSAELAPILETQLNGLFRRTSVLEVE 299  
  
QY 312 NNISTKFGDGGYVYQAIQRPVTRINDESRITVDVEYIIDPVHPVYVRRINFTGNFTQDEVLR 371  
DB 300 QRMKSLGERGYATAQNVNHTFDEQDKTISLDFIVEAGKSYTVRQIRFEGNTSSADSTL 359  
  
QY 372 REMRQLEGALASNOKIQLSRARLMRTGFFKHTVTDTRPVPNSPDQVDVNVFVEEQSGS 431  
DB 360 QEMRQEGAWLSSELVELGLRLDRDTGFESVETKTEAIPGS-DQVDVIYKVKERNITG 418  
  
QY 432 STTAAGYSQSGGVTFQPDVSONNFMGKHKVNASFSETSRETVSLGNTNPFYTVNGVSQ 491  
DB 419 INFGIGYGTESGLSYQASIKQDNFLGMGSSISLGGTRNDYGTITNLGYNPEYFTKDGV-- 476  
  
QY 492 SLSSGYIRKTKYD--NKNISNVLDYSGLSYGYPIDENQRISEGLNADNTKLHGGRFMG 548  
DB 492 SLSSGYIRKTKYD--NKNISNVLDYSGLSYGYPIDENQRISEGLNADNTKLHGGRFMG 548
```

```

Db 477 SLGGNVSEEDSKSNTSAGRTSYGGLTLGPPVNNNSYYLGVGYTNKL----- 530
QY 549 GISNV-KOLMADGGKIQVNDNNGIPDEF-HDYTTYNAILGNWYSSSLDRPVPFTQGMHSHVD 606
Db 531 --KNTAPEVNRDLYRQSMKYNDSWTFKSHD---FDSLFGWVNSLNRGVPFTKGVANIG 585
QY 607 --LTVGFDKTHOKVYVYOGNIYRPF-----IKSVLRGVAKLGYGNLDPFYNF 653
Db 586 GRVTIPGSDNKKYLKNAEAOGFYPLDRHGWVLSRISASFADGFS---GKRLPFYQY 641
QY 654 YAGGSGVGVGDOSSLGPRSQAYLTARRGQQTLL--GEVVGNALATGSELILPLPKG 711
Db 642 SAGGSGVLRGFRAYGAIGPNA-IYTRQCPDSVCLVSSDVIGGNAMVTASTELIVPTPEVA 700
QY 712 DW-IDQVRPVIIEGGQVFDTTGMDKQITDLTQFKDPQATAPQNAKAA--NRPLLTQDQK 768
Db 701 DKNQSVRTSLFVDAASVWNT-----RWKAEDKAKAKLNVDPYSDPSR 744
QY 769 LRSAGVGTATWTPIGPLSISYAKPLNKQNDQDTDTVQFQIGSVF 813
Db 745 VRASAGVALOWQSPICGLVFSYAKPLKRYQGDIEIQFQFISGGTF 789

RESULT 14
Q9ZE03 PRELIMINARY; PRT; 768 AA.
AC Q9ZE03;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMPI (OMPI).
GN RPI160.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=MADRID E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
DR EMBL: AJ235270; CAAL4627.1;
DR InterPro: IPR00184; Bac_surfa9_d15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 768 AA; 87196 MW; BIAB12D783D9DFCB CRC64;

Query Match 17.98; Score 754; DB 16; Length 768;
Best Local Similarity 26.38; Pred. No. 2.le-33;
Matches 216; Conservative 145; Mismatches 383; Indels 76; Gaps 20;

QY 7 KGFQVSAMTAVMMVMSTHAQAADPMANDITITGLQRTVIESLQSVLPFRLGQVYSENL 66
Db 2 KIISIKTLILLIFHYHISFADYVIRKIYIEGNHRVERSTIESYKLNVGTEYNNSE 61
QY 67 ADGVKALYATNFSDVQVY-HQEGRIYQVTERPLIAEINFEGNRLIPKEGLQGLK 125
Db 62 DEAIKRLYATSLFRINNMVITDGNLIVNVTETPTPISWFSGNSKIRTNILAKEIYMS 121
QY 126 LAVQPLQKQAVQVMEIETLNTQYISQGYNTETVQKQTMLDGNRVKLDMTAEKPARV 185
Db 122 ---GESLSQAKIEDVKKILEIYRSRGFSTKVPKIKSLNNRVKVFIDFAEGPKTVIK 178
QY 186 DINTIIGNHFSDDADIDVLAIKDNK-INPLSKADRYTQKLVTSLENRAKYLNAAGVRF 244
Db 179 SIYFSGNEHYSDELKSLVLTRESKRWFRLESNDTDPDRVEYDKELLREFYQSVGFAD 238

```

```

QY 245 EIKDAKLNINEDKNRIFVEISLHSEGEYRFGQTQFLGNLTYYQ-AEALALKKFAEBGFS 303
Db 239 RVISASVALNDTKETITYSIEEGEKYRFGNVTIDNKLTNINIKLNKVINIKQGRIFN 298
QY 304 QAMLEQTTNNISTKFGDDGYYIAIRPVTRINDESRVDEYIDVPVHYVVRINPTGN 363
Db 299 MKTVDDIAEKIGEYFTANGYPAYNVYPDIKKND-NHTADIKFTIEKADKYVYINKININ 357
QY 364 FKTQDVLRREROLGASLNOKIOLSRARLMRTGFFKHVTVDTRPVNPSQDQVDNEV 423
Db 358 LKTEDHIVIRAPKIEGDVWNSYIEKGERNLNLDYFEKVSISLAQT-KAKDKYDVNVE 416
QY 424 VEEQPSGSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNASFSRSRETVYSLGTNTPY 483
Db 417 VDEKSTSSIGFDLGYNTAGLFGFRSFLERNLVGTCKLLNAGVQVSKNSTSYGCTDPH 476
QY 484 FTVNGVSQSLGYY----RKYDKNKINSYVLDSVGGSLSYGYPIDE--NQRISEGLNA 537
Db 477 FLDRDLSLVNAPRNVYTRGASVINTDQSYKLHSIGVKLSIDYDMKEDLSHEIDYLKR 536
QY 538 DNTKLGHRFRMGISNVKQLMADGGKITQVNDNNGIPDFKHVDYTYNAILGWNYSISLDRPVFP 597
Db 537 DILSAPSP-----SNSIFLNEOMGKLITSAG-----HTIT-----YDQDNKIVP 577
QY 598 TQG--MSHVDLTGVGDKTHQKVVYQGNIRPFIKSVLRGVAKLG-----YGNLPPF 649
Db 578 KNGYLVSQTQEPAGVGGDNKYIKHEIECKFYKSFINKVYTKLSAAGDMAGLGGKRVRI 637
QY 650 YENFVAGGVSVRGYDQSSLGPRSQAYLTARRGQQTTLGEVVGCGNALATGSELILPLPF 709
Db 638 SDRFNLDGY-SLRGFASGGVGR-----EKNTEGLEGGERYYTFSTELNFPVPV 685
QY 710 KGDVTDQVRPVIIEGGQVFDTTGMDKQITDLTQFKDPQATAPQNAKARNPLLTQDKOL 769
Db 686 PEEF-NFTGAVFIDLGSVWG-VGLNKK-----QYTPNG-----FYNDQSL 724
QY 770 RYSAGVGATWYTPIGPLSISYAKPLNKQNDQDTDTVQFQI 809
Db 725 RASVGFGEFTWTRFAPIRMWGFPIKKQYD--DTQNFHL 762

RESULT 15
Q92Q48 PRELIMINARY; PRT; 776 AA.
AC Q92Q48;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE PUTATIVE OUTER MEMBRANE TRANSMEMBRANE PROTEIN.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Rasperger U., Surzycki R., Thebault P., Vandenbol M.,
RA Vorhoeiter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR EMBL: AL591787; CAC46081.1; -.
KW Complete proteome.
SQ SEQUENCE 776 AA; 84520 MW; CE261E6372A94726 CRC64;

```

Query Match 17.8%; Score 748.5; DB 16; Length 776;
Best Local Similarity 25.5%; Pred. No. 4.2e-33;
Matches 216; Conservative 152; Mismatches 354; Indels 125; Caps 20;

QY 11 VSAMTAMVMVM-----STHAQAADFMANDITITGLQRYTIESLQSVLPFRLGQV 60
DB 11 VSAPALSASWATGTGVLVAGTSAQA--VINRVEVRGATRVSAETVRANITIVFGKS 68

QY 61 VSENQLADGVKALYATNFSQVYHOEGRIIYQVTERPLIAEINFEGNRLIPKEGLQEG 120
DB 69 FSNADIDASVKRLYATGYSFSDVSTVSGSLVSVSENQLVNVQVFNNGRKIKDDKLGQV 128

QY 121 LKNAGLAVGQPLQATVOMIETELTNOYISQYYINTEITVKTQMLDGNRKLDMTFAEGK 180
DB 129 VRTQPLG---PYSEATVETDQAIIRDAYAAIGRSADVTTTQVVPVIAEGRVNLAFVINEGE 185

QY 181 PARVVDINIIGNOHFSADLDIVLAIKDKNI-NPLSKADRYTQEKLVTSLENLRAKYLNA 239
DB 186 RTKITQINFVGNVEYISGRVQSIVATRESGIFSELTTRDVPNPKLRADBELLRQFYNR 245

QY 240 GFVRFEIKDAKLINEDKNRIFVEISLHEGEQYRFQGTQFLGNLTYYTQA-ELEALLKFKA 298
DB 246 GYADFQVVSSEALNEATNEVTTITIEGPRYDFGPNVIESTVEGIDAEELRGLVQSRE 305

QY 299 BEGFSQAMLEQTTNNISTKFGDDGYIAQIRPVTRINDESKTVDEYIIDPVHPVYVYRRI 358
DB 306 GTVYKAKDIOSTASEISKRVASEGYPARVTPRGNRDLNHTIAVDYLVQGERAYVERI 365

QY 359 NFGTGNKTQDEVLRRERQLEGALASNQKIQLSRARLMRTGFKHVTVDTRPVNSPDQV 418
DB 366 EIRGNTRTDYVIRREFDVGEQDAFNOEMVARAKRLREALGYFSSVNIQTQ-GSAADRV 424

QY 419 DVNFVVEEQSPGSGSTIAAGYS--QSGGVTEQFDVSNFMGTGKHVNASFSRSETREYVS 476
DB 425 VIVVDVQDQSTGFGIGAGYSAGDGGGFLVEASTEEKNFLRGQYIRLAAGKGEDSQTYN 484

QY 477 LGMTNPFYTVNGVSQSLSGYYRKYTKYDNKNISNVVLDSDYSGSLSYGYPIDENORISFGLN 536
DB 485 VSFTPEYFLGYRLAAGFDLEKFNENDEDD--DNYSYNDQGFSLRVATPITENLSTTLRYN 541

QY 537 ADNTKLHGRFMGISNVKQLMADGKIQVDNNGIPDFKHDTYTYNAILGNWYSSLDPRVF 596
DB 542 YTELEFGDR-DELSPPYDRVIDGS-----PWTRSSISQSITYNFLDDAQL 586

QY 597 PTQGMHSVDLT-VGFGDKTHQKVYQGNTRYPPFIKSVLRGYAKLGY-----643
DB 587 PHEGILASVTQEFAGLGTS-----DFYK-----LTGRKAKWYITLHDEADIIIGSL 631

QY 644 -----GNLPPFENFYAGGYSVRGYDQSSSLGPRSQAYLTARRGQOTTLGEVVGG 693
DB 632 SGSAGHLFETSGSLEVPDQFLNS-NDIRGFERNGLGPRMNN-----GDALGG 678

QY 694 NALATFGSELILPLP-----FKGDWIDQVRPVIFIEGGQVFDTTGMDKQTDLTQFKD 746
DB 679 TTYFTASAEATFPPLGPRDSGFRG-----ALFVDACTLYG-----ND 716

QY 747 PQATAEQNAKAAANRPLLTQDKQLRYSAGVGATWYTPIGPLSISYAKPLNKKQNDQTDTVQ 806
DB 717 VEIGPGESVRGDN-----ASLRASVGVSLIWASPEGPLRVDYAVFVAKEDFDEVQNFK 769

QY 807 FQIGSVF 813
DB 770 FGINSF 776

Search completed: September 5, 2002, 10:03:04
Job time: 538 sec

PD 5/3/99

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:56:07 ; Search time 61.46 Seconds
(without alignments)
1469.297 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRSYFKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 15

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_032802.*

- 1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	813	100.0	813	21	AA1980
2	749	92.1	813	21	AA1981
3	16	2.0	17	21	AA1982
4	15	1.8	16	21	AA1983

ALIGNMENTS

RESULT 1
AA1980

ID	AA1980	standard; Protein; 813 AA.
XX	AA1981	
AC	AA1982	
XX	AA1983	
DF	14-MAR-2000	(first entry)
XX	14-MAR-2000	
DE	M. catarrhalis BAS027	polypeptide #1.
XX	BAS027	OMP85; outer membrane protein; otitis media; treatment;
KW	diagnosis; bacterial infection.	
XX	Moraxella catarrhalis.	
OS	WO9963093-A2.	
XX	WO9963093-A2.	
PN	09-DEC-1999.	
XX	09-DEC-1999.	
PF	31-MAY-1999;	99WO-EP03822.
XX	31-MAY-1999;	
PR	03-JUN-1998;	98GB-0011945.
XX	03-JUN-1998;	
PR	08-MAR-1999;	99GB-0005304.
XX	08-MAR-1999;	
PA	(SMIK) SMITHKLINE BEECHAM	BIOLOGICALS.
XX	(SMIK) SMITHKLINE BEECHAM	
PI	Vinals-Bassols C;	
XX	Vinals-Bassols C;	
DR	WPI; 2000-105700/09.	
XX	WPI; 2000-105700/09.	
DR	N-PSDB; AA229550.	
XX	N-PSDB; AA229550.	
PT	Novel BAS027 polynucleotide and polypeptides from Moraxella	
XX	Novel BAS027 polynucleotide and polypeptides from Moraxella	
PT	catarrhalis useful for treating M. catarrhalis infection such as otitis	
XX	catarrhalis useful for treating M. catarrhalis infection such as otitis	
PS	Claim 1; Fig 3; 109pp; English.	
XX	Claim 1; Fig 3; 109pp; English.	
CC	The present sequence is BAS027 polypeptide, which shows significant	
XX	The present sequence is BAS027 polypeptide, which shows significant	
CC	homology to Neisseria meningitidis OMP85 outer membrane protein. It is	
XX	homology to Neisseria meningitidis OMP85 outer membrane protein. It is	
CC	encoded by DNA obtained from chromosomal DNA library of Moraxella	
XX	encoded by DNA obtained from chromosomal DNA library of Moraxella	
CC	catarrhalis strain Mc2931 (ATCC 43617). BAS027 polynucleotide and	
XX	catarrhalis strain Mc2931 (ATCC 43617). BAS027 polynucleotide and	
CC	polypeptide can be used for diagnosis and staging of disease, determining	
XX	polypeptide can be used for diagnosis and staging of disease, determining	
CC	susceptibility to a disease and to prepare medicaments for treating M.	
XX	susceptibility to a disease and to prepare medicaments for treating M.	
CC	catarrhalis infections, especially otitis media. The BAS027 DNA can be	
XX	catarrhalis infections, especially otitis media. The BAS027 DNA can be	
CC	used as probe for screening of genetic mutations, serotype, taxonomic	
XX	used as probe for screening of genetic mutations, serotype, taxonomic	
CC	classification or identification. BAS027 agonists, antagonists and	
XX	classification or identification. BAS027 agonists, antagonists and	
CC	antibodies may be used to prevent and/or treat bacterial infections.	
XX	antibodies may be used to prevent and/or treat bacterial infections.	
SQ	Sequence 813 AA;	
Query Match	100.0%; Score 813; DB 21; Length 813;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches	813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1 MRSYFKGFQVSAMTAVMMVMMSTHAQAADFMANDDITGLQRTVIESLQSVLPFLQGV 60	
Db	1 mrsyfkqfqsamtavmmvmmsthaqaadmfanditglqrvtieslqsvlpflqgv 60	
Qy	61 VSENLADGVKALYATGTFSDVQVYHQBGRVYVTEITVKTMLDGNRKLDMTFAEGK 120	
Db	61 vsenladgvkalyatgtnfsdvqvyhqbgrvlyqterpliaefnegrllpkleglqeg 120	
Qy	121 LKNAGLAVGQPLKQATVOMIEFTLNQYISQGYNYNTEITVKTMLDGNRKLDMTFAEGK 180	
Db	121 lknaglavgqplkqatvqmieteltnqyisqgynteitvktmldgnrvklmdtfaek 180	
Qy	181 PARVVDINIINQHFSDADLDVLAIKDNKINPLSKADRYTQEKLVTSLENIRAKYLNAG 240	
Db	181 parvvdinlignqhfdsadldvlaikdnklnplskadrytqeklvtslenirakyl 240	
Qy	241 FVRFEIKAKLNINEDKNRIFVEISLHEGEQYRFQGTQFLGNLTYYQAEALLKFAEE 300	
Db	241 fvrfeikaklninedknrfveislhegeqyrfqgtqflgnltqaealllkfaee 300	

QY 301 GFSQAMLEQTNNTNISTKFGDGGYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINF 360
 Db 301 gfsqamleqtntnnistkfgdggvyaaqirpvttrindesrtvdveyidpvhpyvrrinf 360
 QY 361 TGNFKTQDEVLRERMQLEGALASQKIQLSRARLMRTGFFKHVTVDTTRVPVNSPDQVDV 420
 Db 361 tgnfktdqdevlrermqlegalasnkqqlsrrarlmrtdgffkhvtdtrvpvnsdpdqv 420
 QY 421 NFVVEQPSGSSTIAAGYSQSGGVTQFDVSONNFMTGKHNASFSRSTRETVYSLGMT 480
 Db 421 nfveeqpsgsstiaagysqsggvtqfdvsnnfmgtkhnasfsrsetrevyslgmt 480
 QY 481 NPYFTVNGVSQSLSGYRRTKYDNKNISNVLDYSYGGSLSYGYPIDENORISFGLNADNT 540
 Db 481 npyftvngvsqslsgyrrtkydnknisnyvldysygsysgyypidenqrtsfglnadnt 540
 QY 541 KLHGGRFMGTSNVKQLMADGKIQVDNNGIPDFKHDTTYNAILGNWYSSLDPRVPPTQG 600
 Db 541 klhggrfmgtisnvkqlmadgkqlvdnngipdfkhdtytnailgnwysldprvpptqg 600
 QY 601 MSHSVDLTVGFGDKTHQVYQGNIRPFTKKSVLRGYAKLGYGNLPPFENFYAGGYGS 660
 Db 601 mshsvdltvfgdkthqvyyqgnirpftkksvlgryaklgygnlppfeyenyaggygs 660
 QY 661 VRGYDSSSLGPRSQAYLTARRGQOTTLGEVVGGNALATFGSELILPLPFKGDWIDQVRPV 720
 Db 661 vrygdssslgprsqayltarrgqottlgevvggnalatfgselilplpfkgdwidqvrpv 720
 QY 721 IFTEGGVFTTGMDKQOTIDLTQFKDQATAEQNAKARNPLLTQDKQLRYSAGVGATWY 780
 Db 721 ifteggvfttgmdkqotidltqfkdpqataeqnakaarnpiltqdkqlrysagvgatwy 780
 QY 781 TPIGPLSISYAKPLNKKNDQTDTPVQFQIGSVF 813
 Db 781 tpigplsisyakplnkkndqtdtpvfqigsvf 813
 RESULT 2
 AAY44391
 ID AAY44391 standard; Protein; 813 AA.
 XX AC AAY44391;
 XX DT 14-MAR-2000 (first entry)
 XX DE M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.
 XX KW BASB027: OMP85; outer membrane protein; otitis media; treatment;
 XX KW diagnosis; bacterial infection.
 XX OS Moraxella catarrhalis.
 XX PN W09963093-A2.
 XX PD 09-DEC-1999.
 XX PF 31-MAY-1999; 99WO-EP03822.
 XX PR 03-JUN-1998; 98GB-0011945.
 XX PR 08-MAR-1999; 99GB-0005304.
 XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX PI Vinals-Bassols C;
 XX DR WPI: 2000-105700/09.
 XX DR N-PSDB; AAZ29551.
 XX PT Novel BASB027 polynucleotide and polypeptides from Moraxella
 XX PT catarrhalis useful for treating M. catarrhalis infection such as otitis
 XX PT media
 PS Claim 1; Page 102-104; 109pp; English.

XX The present sequence is BASB027 polypeptide, which shows significant
 CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is
 CC encoded by DNA obtained from chromosomal DNA library of Moraxella
 CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and
 CC polypeptide can be used for diagnosis and staging of disease, determining
 CC susceptibility to a disease and to prepare medicaments for treating M.
 CC catarrhalis infections, especially otitis media. The BASB027 DNA can be
 CC used as probe for screening of genetic mutations, serotype, taxonomic
 CC classification or identification. BASB027 agonists, antagonists and
 CC antibodies may be used to prevent and/or treat bacterial infections.
 XX Sequence 813 AA;
 SQ
 Query Match 92.1%; Score 749; DB 21; Length 813;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 65 QLADGVKALYATGNFSDVQVYHOGRIIYQVTERPLIAEFNFGNRLIPKEGLOEGLKNA 124
 Db 65 qladgvkalyatgnfsvdvyhqeagriiyyqterpliaefnfgnrlipkeglqeglkna 124
 QY 125 GLAVGQPLKQATVQMIETELTNOYISOGYYNTEITVKQTMLDGNRVKLDMTFAEGKPARV 184
 Db 125 glavgqplkqatvqmieteltdnyisogyyntelvtvkqtdmldgnrvklmdmtfaegkparv 184
 QY 185 VDINIIGNHFSDADLIDVLAIKDNKINPLSKADRYQEKLVTSLENLRAKYLNAGVRF 244
 Db 185 vdiniignqhfdsadlidvlaikdnkinplskadryqeklvtslenlrakylnagvrf 244
 QY 245 EIKDAKUNINEDKNRIFVEISLHEGEQYRFGOTQFLGNLYTQAEALLKFKAEFGSQ 304
 Db 245 eikdakuninedknriifeislhegedyrfqgtqflgnlytqaeallkfkaeefsq 304
 QY 305 AMLEQTTNNISTKFGDGGYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINF 364
 Db 305 amleqttnnistkfgdggvyaaqirpvttrindesrtvdveyidpvhpyvrrinf 364
 QY 365 KTQDEVLRERMQLEGALASQKIQLSRARLMRTGFFKHVTVDTTRVPVNSPDQVDV 424
 Db 365 ktqdevlrermqlegalasnkqqlsrrarlmrtdgffkhvtdtrvpvnsdpdqv 424
 QY 425 EEQPSGSSTIAAGYSQSGGVTQFDVSONNFMTGKHNASFSRSTRETVYSLGNTNPF 484
 Db 425 eeqpsgsstiaagysqsggvtqfdvsnnfmgtkhnasfsrsetrevyslgmtnpf 484
 QY 485 TVNGVSQSLSGYRRTKYDNKNISNVLDYSYGGSLSYGYPIDENQRIISFGLNADNTKLHG 544
 Db 485 tvngvsqslsgyrrtkydnknisnyvldysygsysgyypidenqrissfglnadntklhg 544
 QY 545 GREMGISNVKQLMADGKIQVDNNGIPDFKHDTTYNAILGNWYSSLDPRVPPTQGSMSH 604
 Db 545 grfmgisnvkqlmadgkqlvdnngipdfkhdtytnailgnwysldprvpftqgms 604
 QY 605 VDLTVGFGDKTHQVYQGNIRPFTKKSVLRGYAKLGYGNLPPFENFYAGGYR 664
 Db 605 vdltvfgdkthqvyyqgnirpftkksvlgryaklgygnlppfeyenyaggyr 664
 QY 665 DQSSILGPRSQAYLTARRGQOTTLGEVVGGNALATFGSELILPLPFKGDWIDQVRP 724
 Db 665 dqssilgprsqayltarrgqottlgevvggnalatfgselilplpfkgdwidqvrp 724
 QY 725 GGQVFDFTGMDKQOTIDLTQFKDQATAEQNAKARNPLLTQDKQLRYSAGVGATWY 784
 Db 725 ggqvfdftgmdkqotidltqfkdpqataeqnakaarnpiltqdkqlrysagvgatw 784
 QY 785 PLSTSYAKPLNKKNDQTDTPVQFQIGSVF 813
 Db 785 plstsyakplnkkndqtdtpvfqigsvf 813
 RESULT 3

AAV44393
ID AAY44393 standard; peptide; 17 AA.
XX
AC AAY44393;
DT 14-MAR-2000 (first entry)
XX
DE Moraxella catarrhalis BASB027-specific synthetic peptide #2.
XX
KW BASB027; OMP85; outer membrane protein; otitis media;
KW bacterial infection.
XX
OS Synthetic.
XX
PN WO9963093-A2.
XX
PD 09-DEC-1999.
XX
PF 31-MAY-1999; 99WO-EP03822.
XX
PR 03-JUN-1998; 98GB-0011945.
PR 08-MAR-1999; 99GB-0005304.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Vinals-Bassols C;
XX
DR WPI; 2000-105700/09.
XX
PT Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media -
XX
PS Example 8; Page 64; 109pp; English.
XX
CC The present sequence is a Moraxella catarrhalis BASB027
CC polypeptide-specific peptide which can be coupled to KLH and used for
CC producing antibodies against BASB027. Anti-BASB027 antibodies can be used
CC to treat bacterial infections.
XX
SQ Sequence 17 AA;

Query Match 2.0%; Score 16; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 1e-08;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 676 YLTARRGQQTTLGEVV 691
Db 1 yltarrgqqtllgevv 16
|||||

RESULT 4
AAY44392
ID AAY44392 standard; peptide; 16 AA.
XX
AC AAY44392;
DT 14-MAR-2000 (first entry)
XX
DE Moraxella catarrhalis BASB027-specific synthetic peptide #1.
XX
KW BASB027; OMP85; outer membrane protein; otitis media;
KW bacterial infection.
XX
OS Synthetic.
XX
PN WO9963093-A2.
XX
PD 09-DEC-1999.
XX
PF 31-MAY-1999; 99WO-EP03822.
XX

PR 03-JUN-1998; 98GB-0011945.
PR 08-MAR-1999; 99GB-0005304.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Vinals-Bassols C;
XX
DR WPI; 2000-105700/09.
XX
PT Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media -
XX
PS Example 8; Page 64; 109pp; English.
XX
CC The present sequence is a Moraxella catarrhalis BASB027
CC polypeptide-specific peptide which can be coupled to KLH and used for
CC producing antibodies against BASB027. Anti-BASB027 antibodies can be used
CC to treat bacterial infections.
XX
SQ Sequence 16 AA;

Query Match 1.8%; Score 15; DB 21; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 790 YAKPLNKKQNDQTD 804
Db 2 yakplnkqndqtdt 16
|||||

Search completed: September 5, 2002, 10:04:37
Job time: 510 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:59:52 ; Search time 25.36 seconds
(without alignments)
783.044 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

No matches found

Search completed: September 5, 2002, 10:05:08
Job time: 316 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:03:07 ; Search time 39.77 Seconds
(without alignments)
1964.310 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: September 5, 2002, 10:05:54
Job time: 167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:05:57 ; Search time 20.98 Seconds
(without alignments)
1500.429 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MNSYFKGFQVSAMTMAVM.....LNKKQNDQDTVQFQIGSYF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

No matches found

Search completed: September 5, 2002, 10:12:48
Job time: 411 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:05:12 ; Search time 63.51 Seconds
(without alignments)
2214.531 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRNSYFKGQVSAMTAVMM.....LNKKQNDQTDVQFIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
---------------	-------	----------------	--------	----	----	-------------

No matches found

Search completed: September 5, 2002, 10:12:20
Job time: 428 sec

RESULT	5
C90651	
hypothetical protein ECs0179 [imported]	- Escherichia coli (strain O157:H7, substrain

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:07:32 ; Search time 61.4 seconds
(without alignments)
1470.733 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRSYFKGFQVSAMTMMVMM.....LNKKQNDQTDVTQFQIGSVF 813

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 20

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	813	100.0	813	21	AA1980.DAT
2	749	92.1	813	21	AA1981.DAT

ALIGNMENTS

RESULT 1
ID AAY44390
ID AAY44390 standard; Protein: 813 AA.
XX

AC AAY44390;
XX 14-MAR-2000 (first entry)
XX M. catarrhalis BASB027 polypeptide #1.
XX BASB027; OMP85; outer membrane protein; otitis media; treatment;
KW diagnosis; bacterial infection.
XX Moraxella catarrhalis.
XX W09963093-A2.
XX 09-DEC-1999.
XX 31-MAY-1999; 99WO-EP03822.
XX 03-JUN-1998; 98GB-0011945.
XX 08-MAR-1999; 99GB-0005304.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX Vinals-Bassols C;
XX WPI; 2000-105700/09.
XX N-PSDB; AAZ29550.
XX Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media -
XX Claim 1; Fig 3; 109pp; English.
XX The present sequence is BASB027 polypeptide, which shows significant
CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is
CC encoded by DNA obtained from chromosomal DNA library of Moraxella
CC catarrhalis strain MC2931 (ATCC 43617). BASB027 polynucleotide and
CC polypeptide can be used for diagnosis and staging of disease, determining
CC susceptibility to a disease and to prepare medicaments for treating M.
CC catarrhalis infections, especially otitis media. The BASB027 DNA can be
CC used as probe for screening of genetic mutations. The BASB027 DNA can be
CC classification or identification. BASB027 agonists, serotype, taxonomic
CC antibodies may be used to prevent and/or treat bacterial infections.
XX SQ Sequence 813 AA;
Query Match 100.0%; Score 813; DB 21; Length 813;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRSYFKGFQVSAMTMMVMMVMMSTHAQAADFAMNDITITGLQRTVIESLOSVPRLGQV 60
DB 1 mrsyfkqfqsamtmmvmvmsthaqaadfmndititglqvrtieslqsvprlqgv 60
QY 61 VSENLADGVKALYATGNFSDVQVYHQEGRIYQVTERPLIAEINFEGRNRLPKRGLQEG 120
DB 61 vsenladgvkalyatgnfsvdqvyhqegriyqvtterpliaelnfegnrllpkqglqeg 120
QY 121 LKNAGLAVGQPLKATVOMIETELTNQYISOGYYNTEITVQKTMLDGHRVKLDWTFAGK 180
DB 121 lknaglavgqplkqatvgmieltnqyisggynteitvqktdmldgmrkldwtfagk 180
QY 181 PARVVDINIGNHFSADLDLVLAIKDNKINPLSKADRYTQEKLVTSLENLRAKYLNAG 240
DB 181 parvvdinignhfsadldlvlaikdnkinplskadrytqeklvtslenlrakyl 240
QY 241 FVRFEIKDAKLNEDKNRIFVEISLHGEQYRFQOTFGLNLTYTQAELEALLFKAAE 300
DB 241 fvrfeikdaaklnedknrifeislhgedyrfqotfqlnltytqaeleallfkkee 300
QY 301 GFSQAMLEQTTNNISTKFGDDGYIAQIRPVTRINDESRVTVDVEYIDPVHPVYRRINF 360
DB 301 gfsqamleqttnnistkfgddgyiaqirpvtrindestrvtvdveyidpvhpvyrrinf 360

Db 301 gfsqamleqttnnistkfgddgyvyaqirpvttrindesrtvdveyidvphvpyvrrinf 360
QY 361 TGNFKTQDEVLRREMQLSEALASNOKIQLSRARLMRTGFFKHVTVDTTRPVNPSPOQDV 420
Db 361 tgnfktqdevlrrremrqlsealasngkqlsrlarlmrtgffkhvtdtrpvnpdpqdv 420
QY 421 NFVVEEOPSGSSNTAAAGYSQSGGVTFQFVSONNFMTGKHVNASFSSRSETREVYSILGMT 480
Db 421 nfveeeqpsgstiaagysqsggvtfqfvsqnmfmgtkhvnasfsrsetrevyslgmt 480
QY 481 NPYFTVNGVSQSLSGYRRTKYDNKNISNYVLDYSGGSLSYGYPIDENORISFGLNADNT 540
Db 481 npyftvngvsqslsgyrrtkydnknisnyvldsyggslysgypidengrifsflnadnt 540
QY 541 KLHGGRPMGISNVKQLMADGGKIQVDNNGIPDFKHDTTYTNAILGWNYSLSLDPVPTQG 600
Db 541 klhggrfmgisnvkqlmadggkigvndngipdfkhdyttnailgwnyslsldrpvptqg 600
QY 601 MSHSVDLTVGFGDKTHQKVYVQGNIRPFIKKSVLRGYAKLGYCENNLPPFYENFYAGGYGS 660
Db 601 mshsvdltvfgdkthqkvvyqgnirpfikksvlgayaklgygnnlpfyenfyaggygs 660
QY 661 VRGYDQSSLGPRSQAYLTARRGQOTTLGGEVVGGNALATFGSELILPLPFKGDWIDQVRPV 720
Db 661 vrgydgsslgprsqayltarrgqottlgevgvggnalatfgselilplpfkgdwidqvrpv 720
QY 721 IFTEGGGOVDTTGMCKOTIDLTFQKDPQATAEQNAKAANRPLLTQDKOLRYSAGVGATWY 780
Db 721 iftegggovdttgmckotidltfqkdpqataeqnakaanrp lltqdkqlr y sagvgatwy 780
QY 781 TPIGPLSISYAKPLNKKQNDQTDTRVQFQIGSVF 813
Db 781 tpi gplsisyakplnkkqndqtdtrvqf qigsvf 813
RESULT 2
AAY44391
ID AAY44391 standard; Protein; 813 AA.
XX AAY44391;
AC AAY44391;
XX 14-MAR-2000 (first entry)
DE M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.
XX BASB027; OMP85; outer membrane protein; otitis media; treatment;
KW diagnosis; bacterial infection.
XX Moraxella catarrhalis.
OS WO9963093-A2.
XX PN 09-DEC-1999.
XX PF 31-MAY-1999; 99WO-EP03822.
XX PR 03-JUN-1998; 98GB-0011945.
XX PR 08-MAR-1999; 99GB-0005304.
XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX Vinals-Bassols C;
PI WPI; 2000-105700/09.
XX DR N-PSDB; AAZ29551.
XX Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
XX media
PT
PS Claim 1; Page 102-104; 109pp; English.
XX The present sequence is BASB027 polypeptide, which shows significant

CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is
CC encoded by DNA obtained from chromosomal DNA library of Moraxella
CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and
CC polypeptide can be used for diagnosis and staging of disease, determining
CC susceptibility to a disease and to prepare medicaments for treating M.
CC catarrhalis infections, especially otitis media. The BASB027 DNA can be
CC used as probe for screening of genetic mutations, serotype, taxonomic
CC classification or identification. BASB027 agonists, antagonists, and
CC antibodies may be used to prevent and/or treat bacterial infections.
XX
SQ Sequence 813 AA;
Query Match 92.1%; Score 749; DB 21; Length 813;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 65 QLADGVKALYATGNFSDVQVYHQEGRIYQVTERPLAETNFEGNRLTPKELGELGKNA 124
Db 65 qladgvkalyatgnfsvdvvyhqegril yqvterplaeinfegnrltpk elg elgkna 124
QY 125 GLAVGQPLKOATVQMIETELTNQYISQGYNTETITVKOTMLDGNRVKLDMTFAEGKPARV 184
Db 125 glavgqplk oatvqmieteltnqyisqgynteitvktmldgnrvkldmtfaegkparv 184
QY 185 VDINIIGNHFSDADLIDVLAIKDNKNPLSKADRYTQEKLYTSLENLRKAYLNAGFVRF 244
Db 185 vdiniignhf sda dlidv laikdnknplskadrytqeklytslenlrakaylnagfvrf 244
QY 245 EIKDAKLNINEDKNRIFVEISLHEGEYRFQOTQFLGNLTQTAELALEALLFKAEFGFSQ 304
Db 245 eikdaklninedknri fveislhegeyrfqotqflgnltqtaeleallfk aeefgsq 304
QY 305 AMLEQTTNNISTKFGDDGYYYAQIRPVTRINDESRVTVDVEYYIDPVHPVYVRRINFTGNF 364
Db 305 amleq ttnnistkfgddgyyyaqirpvt rindesrtvdveyyidpvhpvyvrrinf tgnf 364
QY 365 KTQDEVLRREMROJLEGALASNOKIQLSRARLMRTGFFKHVTVDTTRPVNPSPOQDVNFV 424
Db 365 ktqdevlrrremrj legalasnok iqlsrlarlmrtgffkhvtdtrpvnpdpqdvnf v 424
QY 425 EEOQPSGSTIAAGYSQSGGVTFQFVSONNFMTGKHVNASFSSRSETREVYSILGMTNPFY 484
Db 425 eeqpsgstiaagysqsggvtfqfvsqnmfmgtkhvnasfsrsetrevysilgmt npyf 484
QY 485 TVNGVQSLSGYRRTKYDNKNISNYVLDYSGGSLSYGYPIDENORISFGLNADTKLHG 544
Db 485 tvngvqs lsgyrrtkydnknisnyvldsyggslysgypidendr isf glnadtkl hg 544
QY 545 GFEWGISNVKQIMADGGKIQVDNNGIPDFKHDTTYTNAILGWNYSLSLDPVPTQGMSSH 604
Db 545 grfmgisnvkqlmadggkiqvdngipdfkhdyttnailgwnyslsldrpvftqgmshs 604
QY 605 VDLTVGFGDKTHQKVYVQGNIRPFIKKSVLRGYAKLGYCENNLPPFYENFYAGGYGSVRGY 664
Db 605 vdltvfgdkthqkvvyqgnirpfikksvlgayaklgygnnlpfyenfyaggygsvrgy 664
QY 665 DQSSLGPRSQAYLTARRGQOTTLGGEVVGGNALATFGSELILPLPFKGDWIDQVRPVIFTE 724
Db 665 dqsslgprsqayltarrgqottlgevgvggnalatfgselilplpfkgdwidqvrpvifte 724
QY 725 GGOVFDTTGMCKOTIDLTFQKDPQATAEQNAKAANRPLLTQDKOLRYSAGVGATWYTPIG 784
Db 725 ggovfdttgmckotidltfqkdpqataeqnakaanrp lltqdkqlr y sagvgatwyt pig 784
QY 785 PLISISYAKPLNKKQNDQTDTRVQFQIGSVF 813
Db 785 plisisyakplnkkqndqtdtrvqf qigsvf 813
Search completed: September 5, 2002, 10:13:56
Job time: 384 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:09:42 ; Search time 25.27 Seconds
(without alignments)
785.833 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTMAVMM.....LNKKQNDQDTVQFOIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	DB	ID

Description					

No matches found

Search completed: September 5, 2002, 10:14:27
Job time: 285 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:11:12 ; Search time 39.83 Seconds
(without alignments)
1961.351 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTMAVM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

Search completed: September 5, 2002, 10:15:14
Job time: 242 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:14:33 ; Search time 20.99 Seconds
(without alignments)
1499.714 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRNSYFKGFOVSAMTAVMM.....LNKKNDQDTDTVQFIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: September 5, 2002, 10:22:07
Job time: 454 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:14:03 ; Search time 63.52 Seconds
(without alignments)
2214.183 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 813
Sequence: 1 MRNSYKGFQVSAMTMYMM.....LNKKONDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

- Database : SPTREMBL_l9:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp Vertebrate:*
 - 14: sp Unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	DB ID Description

No matches found

Search completed: September 5, 2002, 10:21:40
Job time: 457 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:46:41 ; Search time 88.9 Seconds
(without alignments)
1015.782 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 4202
Sequence: 1 MRNSYFKGFQVSAMTMM.....LNKKQNDQDTVOFGISVF 813

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4202	100.0	813	21	AA1980
2	4189	99.7	813	21	AA1981
3	1123.5	26.7	792	22	AAU03958
4	1123.5	26.7	792	22	AA1984
5	1123.5	26.7	792	22	AA1985
6	1115	26.5	797	22	AAU03957
7	1115	26.5	797	22	AA1984
8	1115	26.5	797	22	AA1984
9	1111	26.4	797	22	AAU03959
10	1111	26.4	797	22	AAU04451
11	1111	26.4	797	22	AA1984

12	1111	26.4	797	22	AA1984
13	1109	26.4	797	21	AA1984
14	1104.5	26.3	792	21	AA1984
15	1052.5	25.0	797	15	AA1984
16	1051.5	25.0	797	15	AA1984
17	1050.5	25.0	797	15	AA1984
18	1047.5	24.9	793	15	AA1984
19	1042	24.8	792	15	AA1984
20	901	21.4	639	22	AA1984
21	569	13.5	896	22	AA1984
22	434	10.3	916	22	AA1984
23	433	10.3	847	19	AA1984
24	429	10.2	925	18	AA1984
25	429	10.2	925	20	AA1984
26	395	9.4	795	20	AA1984
27	393	9.4	787	21	AA1984
28	386.5	9.2	797	20	AA1984
29	379.5	9.0	899	20	AA1984
30	378.5	9.0	891	20	AA1984
31	356.5	8.5	837	20	AA1984
32	348.5	8.3	812	20	AA1984
33	344.5	8.2	614	18	AA1984
34	311	7.4	1249	22	AA1984
35	256.5	6.1	578	21	AA1984
36	249.5	5.9	576	21	AA1984
37	198.5	4.7	469	21	AA1984
38	197.5	4.7	469	21	AA1984
39	197.5	4.7	469	22	AA1984
40	180.5	4.3	624	21	AA1984
41	180.5	4.3	643	21	AA1984
42	180.5	4.3	818	21	AA1984
43	176.5	4.2	919	21	AA1984
44	176.5	4.2	1092	19	AA1984
45	175.5	4.2	889	21	AA1984

ALIGNMENTS

RESULT 1	
AA1984	
ID	AA1984390 standard; Protein; 813 AA.
AC	AA1984390
XX	14-MAR-2000 (first entry)
DT	
XX	M. catarrhalis BAS027 polypeptide #1.
DE	BAS027; OMP85; outer membrane protein; otitis media; treatment;
XX	diagnosis; bacterial infection.
KW	
XX	Moraxella catarrhalis.
OS	
XX	WO9963093-A2.
PN	
XX	09-DEC-1999.
PD	
XX	31-MAY-1999; 99WO-EP03822.
PF	
XX	03-JUN-1998; 98GB-0011945.
PR	
XX	08-MAR-1999; 99GB-0005304.
PP	
XX	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA	
XX	Vinals-Bassols C;
PI	
XX	WPI; 2000-105700/09.
DR	
XX	N-PSDB; AA229550.
XX	
PT	Novel BAS027 polynucleotide and polypeptides from Moraxella
PT	catarrhalis useful for treating M. catarrhalis infection such as otitis
PT	media

Db	181		parvvdlnlgnhfsdadldwlaikdnknplskadrytqeklvtslenrakylnag	240			
QY	241	FVFEIKDAKLNINEDKNRIFVEISLHGEQVRFQGTQFLGNLTVQAELEALLFKFAEE	300				
Db	241	fvfveikdaklninedknrfveislhegeqrfqgtqflgnlntvqaeleallfkfaee	300				
QY	301	GFSQAMLEQTTNNISYKFGDGGYVYQAQIRPVTRINDESRTVDVEYIDVPVHPVYVRRINF	360				
Db	301	gfsqamleqttnnisckfddgyvyaqirpvtrindesrtvdveyyidvpvhpvyrrinf	360				
QY	361	TGNFKTQDEVLRREMQLGECALASNOKIQLSRARLMRTGFFKHVTVDTRPVPSNPDQVDV	420				
Db	361	tgntkqtdevlrrremqrlgealasndkqiglsrarlmrtgffkhvtvdtprvpnsdpqvdv	420				
QY	421	NFVVEEQSGSSITAAAGYSQSGGVTFQFDVSONNFMGTCKGHYNASFSRSETREVYSLGMT	480				
Db	421	nfveeqsgsstiaagysqsggvtfqfdvsqnnfmgtgkhvnasfsrsetrevyslgmt	480				
QY	481	NPYFTVNGVSQSLSGYRRTKYDNNKKNISYVLDYSGGSLSYGYPIDENORISFGLNADNT	540				
Db	481	npftvngvsqslsgyyrktkydnknisnyldsyggslysgypidenqrisfglnadnt	540				
QY	541	KLHGGRFMGISNVKQLMADGGKIQVDNNGIPDFKHDYTTYNAILGWNYSLSLDRPVPTQG	600				
Db	541	klhggrfmglisnvkqlmadggkqlvdnngipdfkhdtytynailgwnyslsldrpvptqg	600				
QY	601	MSHSVDLTVGFGDKTHQKVYVQGNIVRPIKKSVLRCYAKLGVGNLPPYENFYAGGYGS	660				
Db	601	mshsvdltvvgfkdthqkvvyqgnivrpikksvlrcyaklvgynlppfyenfyaggygs	660				
QY	661	VRGYDQSSLGPRSQAYLTARRQQTTLTGEVVGGNALATFCSLILPLPFKGDWIDQVRPV	720				
Db	661	vrygdqsslgprsqayltarrqqttlgevggnalatfcselilplpfkgdwidqvrpv	720				
QY	721	IFTEGGQVDTTCMDKQTTDLTFQKDPQATACONAKAANRPILLTQDKQLURYSAGVCATWY	780				
Db	721	ifteggqvdttdgmkdqtldltfqkdpqataeqnakaanrpilltqdkqlrysaagvcatwy	780				
QY	781	TPIGPLSISYAKPLNKKQNDQTDVTVOFQIGSVF	813				
Db	781	tpigplsisyakplnkkndqtdvtvofqigsvf	813				
RESULT 3							
AAU03958							
ID	AAU03958 standard; Protein; 792 AA.						
XX	AAU03958;						
XX							
DT	23-OCT-2001 (first entry)						
DE	Neisseria meningitidis serogroup A antigenic protein #1.						
XX							
KW	Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell.						
KW	bacterial infection; baculovirus; yeast.						
OS	Neisseria meningitidis.						
XX							
FH	Key	Location/Qualifiers					
FT	Peptide	1..21					
FT		/note= "Signal peptide"					
FT	Protein	22..792					
FT		/note= "Mature N. meningitidis serogroup A antigen"					
XX							
PN	WO200138350-A2.						
XX							
PD	31-MAY-2001.						
XX							
PF	28-NOV-2000; 2000WO-IB01851.						
XX							
PR	29-NOV-1999; 99GB-0028197.						
PR	09-MAR-2000; 2000GB-0005698.						

Db 539 ntynkapkryadfirkygktdg-----adg-----sfkg--llykgtvgwg 577
QY 588 YSSLDPRVPTQGMHSVDLTGVF-GDK-THQKVVYOGNIYRPFPIKKSVLRL-----GYAK 640
Db 578 rnktdsaswptrgyltgvnaeialpgsklqyysathnqtfflpsktftlmlggevagiag 637
QY 641 LGYG--NNLPFYENFYAGGYSGVGYDQSSLGPRSQAYLTARRQQQTTLGGEV--GGNAL 696
Db 638 -gygrtkelpfenfyggglgsvrgyesgtlpgk-----vydeygekisygggnk 686
QY 697 ATFGSELILPLPKGWIDQVRVIFIEGGQVFTDTCMDKQTTIDLTQFKDPQATAEONAK 756
Db 687 anvsaellfmpgskd-artvrlsifadagvw-----dgrtyaaengnnksvysenah 740
QY 757 AANRPLLTQDKQLRYSAGVGTATPPIGLSISYAKPLNKKQNDQTDVTFQFGSVF 813
Db 741 ks-----tftnelrysagavtwsplgpmkfsyayplkkpkpedelqrfqgltgtf 792

RESULT 4
AAB84745
ID AAB84745 standard; Protein; 792 AA.
XX AAB84745;
DT 17-SEP-2001 (first entry)
DE Amino acid sequence of a *Neisseria gonorrhoeae* protein.
KW Serogroup B protein; outer membrane protein; *Neisseria* infection;
vaccine.
XX *Neisseria gonorrhoeae*.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "signal peptide"
FT Protein 22..792
FT /note= "mature protein"
XX WO200152885-A1.
XX 26-JUL-2001.
XX 17-JAN-2001; 2001WO-IB00166.
XX 17-JAN-2000; 2000GB-0001067.
XX 09-MAR-2000; 2000GB-0005699.
XX (CHIR-) CHIRON SPA.
XX Pizza M, Rappuoli R, Giuliani M;
XX WPI: 2001-451895/48.
XX N-PSDB: AAH42129.
XX Composition for treating or preventing infection to, detecting, or for
raising antibodies against *Neisseria* bacteria, comprises an N.
meningitidis serogroup B outer membrane preparation and an immunogenic
component -
XX Disclosure; Page 65-67; 83pp; English.

XX The present sequence represents a *Neisseria gonorrhoeae* protein. The
protein is used to produce the compositions of the invention. The
specification describes a composition, comprising a *Neisseria*
meningitidis serogroup B outer membrane preparation and an immunogenic
component. The immunogenic component is protein disclosed in WO99/57280,
WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
TtpA, TbpA, PilC, OpA, or Omp85. The composition is used for making
a medicament for treating or preventing infection due to *Neisseria*

CC bacteria; a diagnostic reagent for detecting the presence of *Neisseria*
bacteria or of antibodies raised against *Neisseria* bacteria; and/or
CC a reagent which can raise antibodies against *Neisseria* bacteria. It may
CC also be used as a vaccine.
XX
SQ Sequence 792 AA;
Query Match 26.7%; Score 1123.5; DB 22; Length 792;
Best Local Similarity 32.6%; Pred. No. 2.3e-67;
Matches 273; Conservative 153; Mismatches 332; Indels 79; Gaps 19;
QY 10 QVSAMTMVMMVSTHAAQADFMANDITITGLQRTVTSLSQSVLPFRGLQVVSQNLADG 69
Db 2 kllqiasalmlglisplafadftlqdrveglrtepstvfnylpvkvgdyndthgsai 61
QY 70 VKALYATNGSDVOVYHOGRIIVQVTERPLIAEINFEGRNLIPKEGLQELQKAGLAVG 129
Db 62 ikslyatgdfddrvvetadgqllltvierptgslntgkmlqndaklnlesfglaqs 121
QY 130 QPLKQATVQMIETELTNQYISQYYNTEITVQKTMLDGNRVKLDMTFAEGKPARVDINI 189
Db 122 qylnqatlnqavaglkkeylgrgklniqitpkvtklarnrvdiditidegksakitdlf 181
QY 190 IGNOHFSADLIDVIAIKDNKI-NPLSKADRYTOEKLVTSLNLRKAYLNAGFVRETKD 248
Db 182 egnqvysdrklmrqmsltteggitwlrtsdrfdrqkfaqdmekvtdfyngnyfdfrild 241
QY 249 AKLINEDKNRIFVEISLHGEQYRFQGTQFLGNLT-YTQAELEALLKFAEEGFSQAML 307
Db 242 tdiqtneckrtqtkitvhggrfrvkvslgdnepkaleklltmkpgkyerqgm 301
QY 308 EQTNNISTKFGDDGYIAQIRPVTRINDESRTVDVEYYIDPVHPVYVRRINFTGNFKTQ 367
Db 302 tavlgelnrmgsagayseisvqplpnagktvdfvlhiepgkrkiyveihitgnnkr 361
QY 368 DEVLRREROLEGALASNQIKQLSRALMRTEGFKHVTDFRPVSPNDQDVNFVBEQ 427
Db 362 devvrrelrqmesapydtsklqrskervellgyfdnvqfadvplagtpdkvdlmsiter 421
QY 428 PGSSTIAAGYSQSGVYTFQFDVSONNFMGTGKHVNASFSESTREYVSLQMTNPYFTVN 487
Db 422 stgslidsagvwqdtglvmsagvsqdnlfgtgksaalraarskttlngslsftdpyftad 481
QY 488 GVSQSLSGY--YRKT---KYDNKNISNVVLDYSGLSYGYPIDENORISFLGNADNKL 542
Db 482 gvs---lgydiygkafdpkrkastsvkgyktttagggvrmgipvtceydrvnfglaehltv 538
QY 543 HG-----GREFMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTYNAILGNW 587
Db 539 ntynkapkryadfirkygktdg-----adg-----sfkg--llykgtvgwg 577
QY 588 YSSLDPRVPTQGMHSVDLTGVF-GDK-THQKVVYOGNIYRPFPIKKSVLRL-----GYAK 640
Db 578 rnktdsaswptrgyltgvnaeialpgsklqyysathnqtfflpsktftlmlggevagiag 637
QY 641 LGYG--NNLPFYENFYAGGYSGVGYDQSSLGPRSQAYLTARRQQQTTLGGEV--GGNAL 696
Db 638 -gygrtkelpfenfyggglgsvrgyesgtlpgk-----vydeygekisygggnk 686
QY 697 ATFGSELILPLPKGWIDQVRVIFIEGGQVFTDTCMDKQTTIDLTQFKDPQATAEONAK 756
Db 687 anvsaellfmpgskd-artvrlsifadagvw-----dgrtyaaengnnksvysenah 740
QY 757 AANRPLLTQDKQLRYSAGVGTATPPIGLSISYAKPLNKKQNDQTDVTFQFGSVF 813
Db 741 ks-----tftnelrysagavtwsplgpmkfsyayplkkpkpedelqrfqgltgtf 792

RESULT 5
AAB23786
ID AAB23786 standard; Protein; 792 AA.
XX

PI Giuliani MM, Pizza M, Rappuoli R, Holst J;
XX WPI; 2001-381289/40.
DR N-PSDB; AAS07277.
XX Novel 85 kDa antigen from *Neisseria meningitidis* and *Neisseria*
PT gonorrhoeae, useful in the manufacture of a medicament for treating and
PT preventing *Neisseria* bacteria infection -
XX Claim 1; Fig 1; 92pp; English.
PS The sequence represents a *Neisseria meningitidis* serogroup B 85 kDa
XX antigenic protein. *Neisseria meningitidis* colonises the pharynx, causing
CC meningitis and, occasionally, septicaemia in the absence of meningitis.
CC This antigenic protein is useful in the manufacture of a medicament for
CC treating or preventing infection due to *Neisseria* bacteria, such as
CC meningitis and septicaemia. It is also useful as a diagnostic reagent for
CC detecting the presence of *Neisseria* bacteria or antibodies raised against
CC *Neisseria*, and as a reagent for raising the antibodies. The *Neisseria*
CC nucleotide sequences can be expressed in a variety of different
CC expression systems, for example, mammalian cells, baculoviruses, plants,
XX bacteria and yeast.
SQ Sequence 797 AA;

Query Match 26.5%; Score 1115; DB 22; Length 797;
Best Local Similarity 32.5%; Pred. No. 8.8e-67;
Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTAVMMVMTAAQADFMANDITITGLQRTVTSQSLVPLFRIGQVVSNNQLADG 69
DB 2 kkgiasalmmglspialadftqdirveglrtepstvfnylpvkvgdyndthgsai 61
QY 70 VKALYATGNFSDVOYHOGRIITQVTERPLIAEINFGNRLIPKELQGLKKNAGLAVG 129
DB 62 iklsiatgfdvdrvetadgllltvterptglsnltgamlqndaknlesfglaqs 121
QY 130 OPLQATVOMIETETNOYISOGYNTETITVKOTMLDGNRVKLDMTFAEGKPARVVDINI 189
DB 122 qyfqatinqavagkeeylgrglnlqtpkvtklarnrvdditidegksakitdlief 181
QY 190 IGNOHFSADLIDVLAIDKNKI-NPLSKADRYTOEKLVTSLNRAKYLNAGFVRFEKID 248
DB 182 egngvysarklmrgslteggilwtlrsngfneqfagdmekvtdfygnngydfdrild 241
QY 249 AKLNINDEKNRIFVETSLHEGEQVREFGOTQFLNLT-VTQAELEALKKFAEEGFSQAML 307
DB 242 tdiqtnedkktqtkitvheggrfwgkvsiegtnevpkaeleklumkpgkwyergqm 301
QY 308 BOTTNNISTKFGDCGYVYAQRPVTRINDESRVTDVEYIIDVPVYVYRRINFTGNFKTQ 367
DB 302 tavlgelnrmsgagayseisvqlpnaektvdfvliheggrkiyveihitgnnktr 361
QY 368 DEVLREMRQLEGALASNOKLTOLSRALMRGFFKHVTVDTRPVPNSPDQVDVNFVVEEQ 427
DB 362 devvirelrqmesapydsklqrskervellgyfdnvqdvavplagtpdkvdlmmslter 421
QY 428 PSGSSTIAAGYSQSGVTFQFDVSNFMMGTGKHNVSFSRSRETVYSLGNTNPFYTVN 487
DB 422 stgslidlsagvqdtgvlmsagvsqdnlfgtgksaalraskrtklngslsftdpyftad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDLSYGSLSYGYPIDENQRISFGLNADNTKLHGG 545
DB 482 gvslygdyvgkafdrpkastsikyktttagagirmsvpteydevnfglvaehltv-- 539
QY 546 RFMGLSNVQKLMADGGKTOVDNNGIPDPFKHDYTTNAILGNWYSLSLDRPVPVTCQMSHSV 605
DB 540 ---tynkaphkhyadfkkygktgdg-tdgsfkglwykgtvgwgrnkttdsalwptgrlyltgv 595
QY 606 DLTGCVG-GDK-THQKVYVQGNRYRPFIRKSVLR-----GYAKLGYG--NNLPFFYENFYAG 656
DB 596 naeialpagsklyqysathnqtwtfflpslktftlmldgevgiag-gygrtkelpffenfygg 654

QY 657 GYGSVRCYDOSSLGRSQAYLTARRGOQTTLGEVV--GGNALATFGSELILPLPFKGDWI 714
DB 655 glgsrvrgyesgtlqpk-----vydeygekisvgnknkanvsaellfpmgskd-a 703
QY 715 DOVRPVIFIEGGQVDFDTGMDKOTIDLTQFKDPOATAPQNAKAANRPILLTODKQLRYISAG 774
DB 704 rvtlslsfadagsvwdgktyddnssat-----ggrvqniygagnthkstkftneirysag 758
QY 775 VGATWYTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813
DB 759 gavtswlspglpmkfsyaypkkpedelqrfgqlgttf 797
RESULT 7
AAB84744
ID AAB84744 standard; Protein; 797 AA.
XX AAB84744;
XX AC
XX 17-SEP-2001 (first entry)
DT Amino acid sequence of a *Neisseria* serogroup B protein.
DE Serogroup B protein; outer membrane protein; *Neisseria* infection;
XX vaccine.
XX *Neisseria meningitidis*.
OS
FH Key Location/Qualifiers
FT Peptide 1..21 /note= "signal peptide"
FT Protein 22..797 /note= "mature protein"
FT
XX WO200152885-A1.
PN 26-JUL-2001.
PD
XX 17-JAN-2001; 2001WO-IB00166.
XX
XX 17-JAN-2000; 2000GB-0001067.
PR 09-MAR-2000; 2000GB-0005699.
XX
XX (CHIR-) CHIRON SPA.
FA Pizza M, Rappuoli R, Giuliani M;
PI WPI; 2001-451895/48.
DR N-PSDB; AAB42128.
XX
XX Composition for treating or preventing infection to, detecting, or for
PT raising antibodies against *Neisseria* bacteria, comprises an N.
PT meningitidis serogroup B outer membrane preparation and an immunogenic
PT component -
XX
PS Disclosure; Page 59-61; 83pp; English.
XX
CC The present sequence represents a *Neisseria* serogroup B protein. The
CC protein is used to produce the compositions of the invention. The
CC specification describes a composition, comprising a *Neisseria*
CC meningitidis serogroup B outer membrane preparation and an immunogenic
CC component. The immunogenic component is protein disclosed in WO99/57280,
CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
CC WO99/31132, WO99/58683, WO99/58773, and/or N. meningitidis protein PorA,
CC TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making
CC a medicament for treating or preventing infection due to *Neisseria*
CC bacteria; a diagnostic reagent for detecting the presence of *Neisseria*
CC bacteria or of antibodies raised against *Neisseria* bacteria; and/or
CC a reagent which can raise antibodies against *Neisseria* bacteria. It may
CC also be used as a vaccine.
XX Sequence 797 AA;

KW	CG motif; complete Freund's adjuvant; phosphorothioate; immunogenic; Neisseria antigen; Neisseria meningitidis; Neisseria gonorrhoeae; bactericidal; antibacterial; vaccine; immunostimulatory; infection; immune response.																			
	Neisseria meningitidis.																			
	W0200050075-A2.																			
	31-AUG-2000.																			
OS	09-FEB-2000; 2000WO-IB00176.																			
	26-FEB-1999; 99US-0121792.																			
	(CHIR-) CHIRON SPA.																			
	Grandi G, Rappuoli R, Giuliani MM, Pizza M; WPI; 2001-015529/02.																			
XX	Immunogenic composition useful for stimulating an immune response in a mammal against Neisseria infection, comprises Neisseria antigen and an adjuvant composition comprising an oligonucleotide with a CG motif -																			
	Claim 22; Page 32; 39pp; English.																			
	The present invention describes an immunogenic composition (I) comprising a Neisseria antigen and an adjuvant composition comprising an oligonucleotide comprising at least 1 CG motif. Also described is an adjuvant composition (II) comprising an oligonucleotide which comprises at least 1 CG motif and a complete Freund's adjuvant (CFA), where the oligonucleotide preferably comprises at least one phosphorothioate bond. AA92359 to AA92385 represent specifically claimed oligonucleotides of the present invention. (i) is useful for stimulating an immune response in a mammal, preferably a human, against Neisseria infection, preferably Neisseria meningitidis infection and in the manufacture of a medicament for inducing a protective immune response in a mammal. The present sequence represents the claimed Neisseria meningitidis serogroup B amino acid sequence disclosed in GB-9928197.4, which is given in the present invention.																			
	Sequence 797 AA;																			
Query Match 26.5%; Score 1115; DB 22; Length 797;																				
Best Local Similarity 32.5%; Pred. No. 8.8e-67;																				
Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;																				
QY	10	QVSAMTMAVMVMSTHAQAADFMANDDITITGLQRTVTIESLSQSVLPRLQGVSENOLADG	69																	
DB	2	klkqiasalmmlgisplaladftiqdirveglqrtepstvfnylpkvvgdyndthgsai	61																	
QY	70	VKALYATGNFSOVQVYHOGRIIYOVTERPLIAEINFEGRNLIPKEGLQGLKNAGLAVG	129																	
DB	62	iksllyatgffddvrvetadgqllltvterptigslnitgakmlqndaiknlesfglaqs	121																	
QY	130	QPLKQATVQMTELTNQYISOGYNYTEIVKOTMLDGNRVKLDMTFAECPARVVDINI	189																	
DB	122	qyfnqatlnqavaglkveylgrgklniqitpkvtklarnrvdiditdegksakitdief	181																	
QY	190	IGNQHFSDADLDVLAIKDKNI-NPLSKADRYTOEKLVTSLNLRKAYLNAGVFRFEIKD	248																	
DB	182	egnvysdrklmrqmsltteggilwtlrsnqfneqfagdmekvtdfyqngydfdrild	241																	
QY	249	AKLNINEDKNRIEVEISLHGEQYRFGQTQFLGNLT-YTOAELEALLKFKAEFGSQAML	307																	
DB	242	tdiqtnedtktkitvtvheggfrfwgksiegdtnvepkaelkltmkpgkwyerqgm	301																	
QY	308	EQTNNISTKFGDDGYIAQIRPVRINDESRVTVDVEYIDPVHPVYVRRINFTGNFKTO	367																	
DB	302	tavlgelnrmgsagayseisvqplnaetktvdfvlhiepgklyvneihitgnnkr	361																	
QY	368	DEVLRREMOLEGALASNQIKLSRARLMRTGTFKKHVTVDTRPVNSPDQVDVNFVVEEQ	427																	
DB	362	devvirelrmesapydtsklrskervellgyfdnvqfdavplagtpdkvdlmslter	421																	
QY	428	PSGSSTIAAGYSOSGGVTFOFYSNNFMCTGKHVNASFSESTREYVSLGTMNPYETVN	487																	
DB	422	stgldslsagvwdgtgvmagsvsgqnlfgtqksaalrsakttlngslsfcdpyftad	481																	
QY	488	GVSOQLSGYYR--KTKYDNKNISNYLVDSYGGSLSYGYPIDENQRISFGLNADNKLHGG	545																	
DB	482	gvslygydygkafprkastsikqyktttagairmsvpvteydrvnfgivaehltvn--	539																	
QY	546	RFWGISNVKQLMADGGKIQVDNNGIPDFKHDYTYNAILGNWYSSSLDRPVFPQTGSHSV	605																	
DB	540	---tynkaphyadifkkygktgdg-tdgsfkgykgtvgwgrnktksalwptrgyltg	595																	
QY	606	DLTVGF-GDR-THQKVYQGNVYRPFYTKKSVLR-----GYAKLGYG--NNLPYENFYAG	656																	
DB	596	naeiaipgsklyysathnqtffpiskttflmlggevgiag-gygrtkte																		

```
QY 368 DEVLREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNSPDQDVNFVVEEQ 427
Db 362 devvrelrimesapydtskqrskervellgyfdnvqfdaavplagtpdkvdlmnsalter 421
QY 428 PGSSTIAAGYSQSGVTFQFDVSONNFMGTGRKHVNASFSRSETREYISLGMNPNYFTVN 487
Db 422 stgslidsagwvqdtglvmsagvsqdnlfqtkgsaalrasrskttlingslsfcdpyftad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDYSGSLSYGYPIDENQORISFGLNADNTKLHG 545
Db 482 gvslygydygkafdrpkastsikqykttagaglrmsvpteydrvnfglvaehltn-- 539
QY 546 RFGMISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFQGMHSV 605
Db 540 ---tynkaphyadfkkygktdg--tdgsfkgwlykgtvgwgrnktdsalwptrgyltg 595
QY 606 DLTVGF-GDK--THQKVYVQGNIRYRPFIKKSVLRL-----GYAKLGYG--NNLPFENFYAG 656
Db 596 naeialpgsklqyysathnqtwffplsktftlmllggevgiag--gygrtkelpffenyf 654
QY 657 GYGSVRGYDQSSIGPRSOAYLTARRGOQTTLGEVV--GGNALATFGSELILPLPFKGDWI 714
Db 655 glgsrvrgyesgtlqpk-----vydeygekiysygnknkanvsaeellfmpgagk-a 703
QY 715 DOVRPVIFTEGGQVFTTGMQKQIDLTQFKDPOATAEQNAKAAANRPLLTQDKOLRYSAG 774
Db 704 rtrvlsifadagswdtktyddnssat-----ggrvqniygagnthkstkftnelrysag 758
QY 775 VGATWYTPIGPLSISYAKPLNKKNDQTDVQFOIGSVF 813
Db 759 gavtclslplgpmkfsyaplkpkpedelqrfqqlgttf 797

RESULT 9
AAU03959
ID AAU03959 standard; Protein; 797 AA.
XX AC AAU03959;
XX DT 23-OCT-2001 (first entry)
XX DE Neisseria gonorrhoeae antigenic protein.
XX KW Antigenic protein; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast; Neisseria meningitidis.
XX OS Neisseria gonorrhoeae.
XX FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /note= "Signal peptide"
FT Protein 22..797
XX /note= "Mature N. gonorrhoeae antigen"
XX WO200138350-A2.
XX PN 31-MAY-2001.
XX PD 28-NOV-2000; 2000WO-IB01851.
XX PF 29-NOV-1999; 99GB-0028197.
XX PR 09-MAR-2000; 2000GB-0005698.
XX (CHIR-) CHIRON SPA.
XX (STAT-) STATENS INST FOLKEHELSE.
XX GIuliani MM, Pizza M, Rappuoli R, Holst J;
XX WPI; 2001-381289/40.
XX DR N-PSDB; AAS07279.
XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria
XX gonorrhoeae, useful in the manufacture of a medicament for treating and
```

```
PT preventing Neisserial bacteria infection -
XX Claim 1; Page 37-39; 92pp; English.
XX The sequence represents a Neisseria gonorrhoeae 85 kDa antigenic protein.
CC N. gonorrhoeae is closely related to N. meningitidis, which colonises the
CC pharynx, causing meningitis and, occasionally, septicaemia in the absence
CC of meningitis. This antigenic protein is useful in the manufacture of a
CC medicament for treating or preventing infection due to Neisseria
CC bacteria, such as meningitis and septicaemia. It is also useful as a
CC diagnostic reagent for detecting the presence of Neisseria bacteria or
CC antibodies raised against Neisseria, and as a reagent for raising the
CC antibodies. The Neisseria nucleotide sequences can be expressed in a
CC variety of different expression systems, for example, mammalian cells,
CC baculoviruses, plants, bacteria and yeast.
XX Sequence 797 AA;
SQ
Query Match 26.4%; Score 1111; DB 22; Length 797;
Best Local Similarity 32.4%; Pred. No. 1.6e-66;
Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;
QY 10 QVSAMTWAVMVMSTHAQAADFMANDITITGLQRTVITESLQSVLPFRIGVQVVSNNOLADG 69
Db 2 kklqiasalmvlgisplalaodftqdlrvveglqrtepstvfnylpkvvgdyndthgsai 61
QY 70 VKALYATGNFSDVQVYHGEIRIIYQVTERPLIAETINFEGRNLIPKEGLEKKNAGLAVG 129
Db 62 iklsiyatgfdvrvetadgqllltvierptgslngakmlqndaklnlesfglaqs 121
QY 130 QPLKOATVOMIETELTNQYISQGYNTETIVKQTMLDGNRVKLDMTFAEGKPARVVDNI 189
Db 122 qyfnqatlnqavaglkceylgrglniqitpkvtklarnrvdidiidgcsaktldlef 181
QY 190 IGNOHFSADADLIDVLAIKDNKI-NPLSKADRYTOBKLTSLNRAKYLINAGFVFEIKD 248
Db 182 egngvysdrklnrqmsltcggiwtlrsnqfneqfqaqmekvtdfygnngyfdrlld 241
QY 249 AKLNINEDKNRIFVEISLHEGEYRFGQTQFLGNIT-YTOAELEALLFKAEEGSQAML 307
Db 242 tdiqtnedktqtkitvhggrfrwkvsviegdtnvypkaelekillmkpgkwyergqm 301
QY 308 EOTTNNISTKFGDDGYVYVAQIRPVTRINDESTRVDVEYIIDVHPVYVRRINFTGNFKTQ 367
Db 302 tavlgelqnrmsagyaayseisvqplpnaekttvfvhiieprkiyvinehitgnktr 361
QY 368 DEVLREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNSPDQDVNFVVEEQ 427
Db 362 devvrelrimesapydtskqrskervellgyfdnvqfdaavplagtpdkvdlmnsalter 421
QY 428 PGSSTIAAGYSQSGVTFQFDVSONNFMGTGRKHVNASFSRSETREYISLGMNPNYFTVN 487
Db 422 stgslidsagwvqdtglvmsagvsqdnlfqtkgsaalrasrskttlingslsfcdpyftad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNVLDYSGSLSYGYPIDENQORISFGLNADNTKLHG 545
Db 482 gvslygydygkafdrpkastsikqykttagaglrmsvpteydrvnfglvaehltn-- 539
QY 546 RFGMISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDPRVPFQGMHSV 605
Db 540 ---tynkaphyadfkkygktdg--tdgsfkgwlykgtvgwgrnktdsalwptrgyltg 595
QY 606 DLTVGF-GDK--THQKVYVQGNIRYRPFIKKSVLRL-----GYAKLGYG--NNLPFENFYAG 656
Db 596 naeialpgsklqyysathnqtwffplsktftlmllggevgiag--gygrtkelpffenyf 654
QY 657 GYGSVRGYDQSSIGPRSOAYLTARRGOQTTLGEVV--GGNALATFGSELILPLPFKGDWI 714
Db 655 glgsrvrgyesgtlqpk-----vydeygekiysygnknkanvsaeellfmpgagk-a 703
QY 715 DOVRPVIFTEGGQVFTTGMQKQIDLTQFKDPOATAEQNAKAAANRPLLTQDKOLRYSAG 774
Db 704 rtrvlsifadagswdtktyddnssat-----ggrvqniygagnthkstkftnelrysag 758
```


CC oligonucleotide preferably comprises at least one phosphorothioate bond.
CC AAA92359 to AAA92385 represent specifically claimed oligonucleotides of
CC the present invention. (I) is useful for stimulating an immune response
CC in a mammal, preferably a human, against *Neisseria* infection, preferably
CC *Neisseria meningitidis* infection and in the manufacture of a medicament
CC for inducing a protective immune response in a mammal. The present
CC sequence represents the claimed *Neisseria meningitidis* serogroup A amino
CC acid sequence disclosed in GB-9928197.4, which is given in the present
CC invention.

XX Sequence 797 AA;

Query Match 26.4%; Score 1111; DB 22; Length 797;
Best Local Similarity 32.4%; Pred. No. 1.6e-66;
Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTAVMVVSTHQAQADFMADITITGLQRTVIESLSQVLPFRGLQVVSENLADG 69
DB 2 k1qiasalmmigispladftiqdirveg1qrtepstvfnylpkvvgdyndthgsai 61
QY 70 VKALYATGNFSDQVYVHOEGRIYVOTRPLIAEFNFGNRLIPKEGLQEGLNAGLAVG 129
DB 62 klsyafdfdrvretadgqllltvierptigsinitgakmlqndaiknlesf1ags 121
QY 130 QPLKQATVQMIETELTNQYISQYVNTETVTKQTMLDGNRVKLDMTFAEGKPARVVDINI 189
DB 122 qyfnqatinqavag1keeylgrgklniqtpkvtklarnrvdidiiditdegksakitdief 181
QY 190 IGNOHFSADLDIVLAIKDNKI-NPLSKADRYTQEKLVTSLENLRAKYNLAGVFRFEIKD 248
DB 182 egngvysdrklmrqmslttegg1wtlrsngfneqkfaqdmekvtdfyngngyfdfrild 241
QY 249 AKLNINDEKNTFVPSILHEGSOYREFGOTQFLGNLT-YTOAELEALLKFAEGFSQAWL 307
DB 242 tdiqnektktqitkvnegrfrwkvsgstegdcnevvpkaeiekl1cmkpgkwyergqm 301
QY 308 EQTTNNISTKFGDDGYYYAQRVPTTRINDESRTVDVEYIIDPVHPVYVRRNFTGNFKTQ 367
DB 302 tavlgelnrgsagayseisvqplpnaetktvfvhlhepgkrlklyneihitgnktr 361
QY 368 DEVLRENRQLEGALASQKQLSARLMRGFFKHVIVDRPVNSPDQDVNVFVBEQ 427
DB 362 devvrrelrqnasapydtsk1qrskervellgyfdnvqfdayplagtpdkvdlmslter 421
QY 428 PSGSTIAAGYSQSGCVTFQFDVSONNFMGCKHVNASFSRSETREVSLSGMTNPFYFWN 487
DB 422 stgslidsagwvqdtglvmsagvsqdn1fgt9ksaalrasrsktllngslsfddpyftad 481
QY 488 GVSQSLSGYR--KTKYDNKNISYVLDSSYGSLSYGYPIDENQRISFGLNADNPKLHGG 545
DB 482 gvslygydygkafdrprkastsikyktttagagirmsvpteydrvnfglvaehltvn-- 539
QY 546 RFMGISNVKQLMADGCKTQVDNNGIPDPFKHDYTTNAILGNWYSSLDPRVPPTQGMHSV 605
DB 540 ---tynkaphkaydfikkyktgdg--tdgsfkglwlykgtvgwgrnktdsalwptrgyltg 595
QY 606 DLTGCF-GDK-THQKVYVQGNIRYRPFIKSVLR-----GYAKLGVG--NNLPFENFENYAG 656
DB 596 naeialpsk1qysatnqtwtfflksftlmlggevgiag-gygrtkeipffenfeygg 654
QY 657 GYGSVGRYDQSSLPSPSOAYLTARRGQQTTLGEEVY--GGNALTFRGSELIILPLPFKGDWI 714
DB 655 g1gsrvygesglgpk-----vydeygekiysygnkkanvsae1llfmpgakt-a 703
QY 715 DQVRPVIFIEGQVDDTTCMDKQTFIDLTQFDPQATQACNAKAAANRPLLTQDKQLRYAG 774
DB 704 rtrvls1fadagsvwdgktyddnssat-----gdrvniqiyagatnkhstftne1ryasg 758
QY 775 VGATWYTPIGPLSISYAKPLNKKQNDQTDVQFOIGSVF 813
DB 759 gavtws1p9pmkfsyayplkxkpede1qrfqf1gttf 797

RESULT 13

AA84947
ID AA84947 standard; Protein; 797 AA.

XX AA84947;

XX 21-AUG-2000 (first entry)

XX Amino acid sequence of outer membrane protein (omp) 85.

XX Outer membrane protein; omp85; gonococcal infection; symptomatic disease;
KW meningococcal infection; protective immune response; vaccine.

XX *Neisseria meningitidis*.

XX WO200023595-A1.

XX 27-APR-2000.

XX 22-OCT-1998; 98WO-US22352.

XX 22-OCT-1998; 98WO-US22352.

XX (UYMO-) UNIV MONTANA.

XX Judd RC, Manning SD;

XX WPI; 2000-339694/29.

XX N-PSDB; AA15156.

XX New isolated outer membrane protein 85 of *Neisseria gonorrhoeae* and N.
PT meningitidis useful for vaccine, therapeutic and diagnostic
XX compositions for gonococcal or meningococcal infections -

XX Claim 41; Page 89-92; 98pp; English.

XX The present sequence represents an outer membrane protein (omp) 85 of
CC *Neisseria meningitidis*. The omp polypeptides and polynucleotides are
CC useful in compositions for use in the prevention, treatment and diagnosis
CC of non-symptomatic gonococcal infection or meningococcal infection and
CC symptomatic disease. They are also useful for the detection of
CC hydriolisation complexes. Antigens and antibodies specific omp proteins
CC also provide diagnostic, therapeutic and prophylactic compositions for
CC the treatment or prevention of the infections described above. The
CC antibodies are useful for inducing a protective immune response in
CC humans or animals with N. gonorrhoeae, N. meningitidis, or other
CC *Neisseria* species. The proteins, antibodies and polynucleotide
CC sequences of the present invention may also be used in the screening
CC and development of chemical compounds such as drugs or vaccines.

XX Sequence 797 AA;

Query Match 26.4%; Score 1109; DB 21; Length 797;

Best Local Similarity 32.4%; Pred. No. 2.2e-66;

Matches 265; Conservative 154; Mismatches 362; Indels 38; Gaps 14;

QY 10 QVSAMTAVMVVSTHQAQADFMADITITGLQRTVIESLSQVLPFRGLQVVSENLADG 69

DB 2 k1qiasalmmigispladftiqdirveg1qrtepstvfnylpkvvgdyndthgsai 61

QY 70 VKALYATGNFSDQVYVHOEGRIYVOTRPLIAEFNFGNRLIPKEGLQEGLNAGLAVG 129

DB 62 klsyafdfdrvretadgqllltvierptigsinitgakmlqndaiknlesf1ags 121

QY 130 QPLKQATVQMIETELTNQYISQYVNTETVTKQTMLDGNRVKLDMTFAEGKPARVVDINI 189

DB 122 qyfnqatinqavag1keeylgrgklniqtpkvtklarnrvdidiiditdegksakitdief 181

QY 190 IGNOHFSADLDIVLAIKDNKI-NPLSKADRYTQEKLVTSLENLRAKYNLAGVFRFEIKD 248

DB 182 egngvysdrklmrqmslttegg1wtlrsngfneqkfaqdmekvtdfyngngyfdfrild 241

```
QY 249 AKLNEDKRIEVEISLHEGEQYREGQTFGLNLT-YTOAELEALLKPKAEEGFSQAML 307
Db 242 tdiqnedkcktkicvheggrfrwgvksiegtnevpkaeleklmtkpgkweryeqm 301
QY 308 EQTNNISTKFGDDGYIAQIRPVTRINDSRVDVEYYIDPVHVPYVVRINFTEGNFKQ 367
Db 302 tavlgeiqnrmgsagayseisvqplpnaetktvdfvhlieprklyvneihitgnnkr 361
QY 368 DEVLRRMRQLEGALASNQIKLSRARLMTGFFKHVTVTRVPNSPDQDVNFVVEEQ 427
Db 362 devrrlrmqesapydtsklrskervellgyfndvqfdaavplagtpdkvdlmslter 421
QY 428 PSGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNASFSEKSTREYVSLGTMNPYFTVN 487
Db 422 stgsldisagvwdqtlvmsagvsqdnlfgtgksaalrasrskttlngsfdfpyftad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNYLDSYGGSLSYGYPIDENORISFGLNADNTKLHG 545
Db 482 gvsldisagvwdqtlvmsagvsqdnlfgtgksaalrasrskttlngsfdfpyftad 481
QY 546 RFGWISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLSLDRPVPEPTQGMHSV 605
Db 540 ---tynkapkryadfklygktdg--tdgskfwlykgtvgwgrnktksalwptgrlytg 595
QY 606 DLTVGF-GDK--THQKVYVYQGIYRPFIFKKSVLK----GYAKLGYG--NNLPFYENFYAG 656
Db 596 naeialpgsklyqysathnqtwtfflpskttfmlggevgiag-gygrtkelpffenyf 654
QY 657 GYGSVRGYDQSSLGPRSOAYLTARRGOQTLGEVV--GGNALATFGSELILPLPKGDWI 714
Db 655 glgsvryesgltgpk-----vydeygekisyggnkkanvsaelifmpgpkd-a 703
QY 715 DQVRPVIIEGQVDFDTGMDKQIDLTQFKDQATAEQNAKAAANRPLLTDQKOLRYSA 774
Db 704 rtvlslsfadagsvwdktyddnssat-----ggrvqnygagntkstkftneirysag 758
QY 775 VGATWYTPIGPLSISYAKPLNKKNDQDVTQFOIGSVF 813
Db 759 gavlwtspgmkfrayaplkpkpedelqrfqfqlgttf 797

RESULT 14
ID AAY84946
XX AAY84946 standard; Protein; 792 AA.
AC AAY84946;
XX
DT 21-AUG-2000 (first entry)
XX
XX Amino acid sequence of outer membrane protein (omp) 85.
XX Outer membrane protein; omp85; gonococcal infection; symptomatic disease;
XX meningococcal infection; protective immune response; vaccine.
XX Neisseria gonorrhoea.
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX /note= "signal peptide"
XX
XX WO200023595-A1.
XX
XX 27-APR-2000.
PD
XX
XX 22-OCT-1998; 98WO-US22352.
PF
XX
XX 22-OCT-1998; 98WO-US22352.
PR
XX
XX (UYMO-) UNIV MONTANA.
PA
XX
XX Judd RC, Manning SD;
PI
XX
XX
```

WPI; 2000-339694/29.
N-PSDB; AAA15155.

DR New isolated outer membrane protein 85 of *Neisseria gonorrhoeae* and *N.*
XX meningitidis useful for vaccine, therapeutic and diagnostic
PT compositions for gonococcal or meningococcal infections
PS
XX Claim 1; Fig 2; 98pp; English.

XX The present sequence represents an outer membrane protein (omp) 85 of
CC *Neisseria gonorrhoeae*. The omp polypeptides and polynucleotides are
CC useful in compositions for use in the prevention, treatment and diagnosis
CC of non-symptomatic gonococcal infection or meningococcal infection and
CC symptomatic disease. They are also useful for the detection of
CC hybridisation complexes. Antigens and antibodies specific omp proteins
CC also provide diagnostic, therapeutic and prophylactic compositions for
CC the treatment or prevention of the infections described above. The
CC antibodies are useful for inducing a protective immune response in
CC humans or animals with *N. gonorrhoeae*, *N. meningitidis*, or other
CC *Neisseria* species. The proteins, antibodies and polynucleotide
CC sequences of the present invention may also be used in the screening
CC and development of chemical compounds such as drugs or vaccines.

XX Sequence 792 AA;

Query Match 26.3%; Score 1104.5; DB 21; Length 792;
Best Local Similarity 32.6%; Pred. No. 4.4e-66;
Matches 268; Conservative 155; Mismatches 350; Indels 49; Gaps 17;

```
QY 10 QVSAMTAVMVMVMTTHAQADFMANDITITGLQRTTESLQSLVLPFRGLQGVVSENLADG 69
Db 2 kklqasalmmlglspafadftlqdrveglrtepstvfnylpvkgvdyndthgsal 61
QY 70 VKALYATGNFSDVOVYHOEGRIIYQVTERPLIAEINFGNRLIPKEGLEKKNAGLAVG 129
Db 62 ikxiyatgdfddrvvetadgillltivcptgislmgakmlqndaklnlesfglqas 121
QY 130 QPLKOATVOMTETELTNQYISQGYNTETVKQTMLDGNRVKLDMTFAEGPARVDINI 189
Db 122 qyfnqatlngavaglkceylgrgklnlqtpkvtklarnrvdltidigesaktidlef 181
QY 190 IGNOHFSADLIDVLAIKDNKI-NPLSKADRYTOBKLTSLNLRKALYNAGVFKEIKD 248
Db 182 egnvqysdrklnmqslceggilwtlfrsdrfdkqfadmekvtdfygnngyffdrild 241
QY 249 AKLNEDKRIEVEISLHEGEQYREGQTFGLNLT-YTOAELEALLKPKAEEGFSQAML 307
Db 242 tdiqnedkcktkicvheggrfrwgvksiegtnevpkaeleklmtkpgkweryeqm 301
QY 308 EQTNNISTKFGDDGYIAQIRPVTRINDSRVDVEYYIDPVHVPYVVRINFTEGNFKQ 367
Db 302 tavlgeiqnrmgsagayseisvqplpnaetktvdfvhlieprklyvneihitgnnkr 361
QY 368 DEVLRRMRQLEGALASNQIKLSRARLMTGFFKHVTVTRVPNSPDQDVNFVVEEQ 427
Db 362 devrrlrmqesapydtsklrskervellgyfndvqfdaavplagtpdkvdlmslter 421
QY 428 PSGSSTIAAGYSQSGVTFQFDVSONNFMGTGKHVNASFSEKSTREYVSLGTMNPYFTVN 487
Db 422 stgsldisagvwdqtlvmsagvsqdnlfgtgksaalrasrskttlngsfdfpyftad 481
QY 488 GVSQSLSGYR--KTKYDNKNISNYLDSYGGSLSYGYPIDENORISFGLNADNTKL 542
Db 482 gvsldisagvwdqtlvmsagvsqdnlfgtgksaalrasrskttlngsfdfpyftad 481
QY 543 HGGREMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSLSLDRPVPEPTQGM 602
Db 539 n-----tynkapkryadfklygktdg--adgsfkglykgtvgwgrnktksalwptgrly 592
QY 603 HSDVLTGVF-GDK--THQKVYVYQGIYRPFIFKKSVLK----GYAKLGYG--NNLPFYENF 653
Db 593 tgvnaeialpgsklyqysathnqtwtfflpskttfmlggevgiag-gygrtkelpffenyf 651
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:47:21 ; Search time 37.39 Seconds
(without alignments)
531.105 Million cell updates/sec

Title: US-09-701-711-2

Perfect score: 4202

Sequence: 1 MNSYFKGFQVSAMTAVNM.....LNKQNDQDTVQFQIGSVF 813

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_patents_AA.*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1052.5	25.0	797	3	US-08-433-522A-2
2	1052.5	25.0	797	3	US-08-433-522A-4
3	1052.5	25.0	797	3	US-08-433-522A-6
4	1052.5	25.0	797	3	US-09-135-166-2
5	1052.5	25.0	797	3	US-09-135-166-4
6	1052.5	25.0	797	3	US-09-135-166-6
7	1052.5	25.0	797	3	US-08-942-046-2
8	1052.5	25.0	797	4	US-08-942-046-4
9	1052.5	25.0	797	4	US-08-942-046-6
10	1047.5	24.9	793	3	US-08-433-522A-10
11	1047.5	24.9	793	3	US-09-135-166-10
12	1047.5	24.9	793	4	US-08-942-046-10
13	1042.5	24.8	792	3	US-08-433-522A-8
14	1042.5	24.8	792	3	US-09-135-166-8
15	1042.5	24.8	792	4	US-08-942-046-8
16	146.5	3.5	2123	4	US-08-968-685A-10
17	142.5	3.4	2199	5	PCT-US95-11684-2
18	141.5	3.4	1599	2	US-08-617-697-9
19	138.5	3.3	1536	1	US-08-038-682-2
20	138.5	3.3	1536	1	US-08-302-832-2
21	138.5	3.3	1536	2	US-08-530-198-2
22	138.5	3.3	1536	2	US-08-469-880-2
23	138.5	3.3	1536	2	US-08-728-470-2
24	138.5	3.3	1536	2	US-08-617-697-2
25	138.5	3.3	1536	4	US-08-719-841-2
26	136.5	3.2	2314	4	US-09-268-347-49
27	136.5	3.2	990	4	US-09-627-376-7

28	134.5	3.2	1178	1	US-08-446-486-5	Sequence 5, Appli
29	134.5	3.2	1178	1	US-08-463-308-5	Sequence 34, Appli
30	134.5	3.2	1182	1	US-08-349-867-34	Sequence 5, Appli
31	134.5	3.2	1182	2	US-08-598-305A-34	Sequence 34, Appli
32	134.5	3.2	1188	1	US-08-239-476-34	Sequence 34, Appli
33	134.5	3.2	1188	2	US-08-639-923A-34	Sequence 34, Appli
34	134.5	3.2	1188	5	PCT-US95-05431-34	Sequence 34, Appli
35	134.5	3.2	2048	4	US-09-268-347-48	Sequence 48, Appli
36	132	3.1	905	4	US-09-074-658-70	Sequence 70, Appli
37	130	3.1	969	1	US-07-671-817A-4	Sequence 4, Appli
38	130	3.1	1177	1	US-07-828-788A-8	Sequence 8, Appli
39	130	3.1	1177	1	US-07-920-085-2	Sequence 2, Appli
40	130	3.1	1177	5	PCT-US92-11337-8	Sequence 8, Appli
41	130	3.1	1177	6	5169629-2	Patent No. 5169629
42	129.5	3.1	2089	1	US-08-418-893D-23	Sequence 23, Appli
43	129.5	3.1	2089	1	US-08-418-893D-24	Sequence 24, Appli
44	128.5	3.1	682	3	US-08-613-009A-10	Sequence 10, Appli
45	128.5	3.1	702	2	US-08-867-941-25	Sequence 25, Appli

ALIGNMENTS

RESULT 1
US-08-433-522A-2
: Sequence 2, Application US/08433522A
: Patent No. 6013514
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/433,522A
: FILING DATE: 12-SEP-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-434 MIS:j.b
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 797 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-433-522A-2

Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;
Oy 23 STHQAADFAMNDITITGLQRTVIESLOSGLPFRIGQVYSENQADGVKALYATGNFSDV 82

```
Db 14 TTTVFAAPFAKDIRVDGVGDLEQOIRASLPVRACORVTDNDVANIVRSLSVSGRFDV 73
QY 83 QVYHOGRI-IYQVTERPLIAEFNFGNRLIPKEGLQGLKNAGLAVGQPLKQATVOMIE 141
Db 74 KA-HQEGDLVWSVAKSIISDVKIKGNSVITPEALKQNDLNGKFKYGVLDLIREKNEFA 132
QY 142 TELTNQYISQGYNTTEITVQOTMLDGNRVKLDMTFAEGKPARVVDINIIGNQHSFADLI 201
Db 133 KSVKHYASVGRYNATPEIVNTLPNNRAELIQINEDDRKAKLASLTFKGNESVSSTLQ 192
QY 202 DVLAIKDNKINPLSKADRYTQELVTSLENLRAKYNAGVFRFEIKDAKLINEDEKKNRIF 261
Db 193 EQMELQPSWKKL-WGNKFEQAQFKDQSIIRDYLYNNGYAKAQITKTDVQNDKTKVN 251
QY 262 VEISLHEGEQYRFQOTFLGNLTQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLOYSARIIIGNLGMSAELEPLLSALHLNDFRSDIADVENAIKALGE 311
QY 321 DGYYAQRVTRINDESRTVDVEYIDPVHPVYVRINFTGNFKTQDEVLRREMRLQEG 380
Db 312 RYGSATVNSVPDFFDANKLTAITLVVDAGRRLTVRQLRFGNTVSADSLRQEMRQEG 371
QY 381 ALASNQIKLSRARLMRTGFKHVTVDTR--PVPNSPDQDVNVFVVEQPSGSSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTPE--TVENRIDPINGSNDEVDVYKVKERNITGSINFGIGY 429
QY 439 SOSGGVTFQFDVSONNFMGTGKHVNASFRRSETREYVSLGMTNPNYTVNGVCSLSGYR 498
Db 430 GTESGISYQASVKQDNFLGTGAASIACTKNDYGTSVNLGYTEPYFTKDG--SLGGNVF 487
QY 499 KTKYDNK---NISNVLDSDYSGSLSYGYPIDENORISFGL-----NADNTKLHGGRFMI 550
Db 488 FENDNKSDDTSNRYKRTYGSNLTGFPVNNNSYVGLGHTYKNKISNPALEYNRNLYI 547
QY 551 SNVKQLMADGKGIQVNDNGIPDFKHDTYTNAILGNWYSSLDPRVFPQOGMSHVD--LT 608
Db 548 QSMK-----FKGNGIKTNDDFS-----FGWYNSLNRGYFTPKGVKASLGGGRVT 592
QY 609 VGFGDKTHOKVYVQGNIRPE-----IKKSVLRGYAKLGYGN-NLPFFENFVAGGYGS 660
Db 593 IPGSDNKYKLSADVOGYPFLDRHLWVVSASAKAGYAN-GFGNKRFLFYQTYTAGGIGS 651
QY 661 VRGYDQSSLSGRSQAYLPAHQOQTLG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGEAYGSGIPNA---IYAEYCGNGSGTGFKKIISDVIIGNAIATASAEILVPTFPVSDK 708
QY 714 I-DQVRPIFIEGGQVFDTT-GMDKQITDLTQFKDPQATAEQNAKAAANRPLIITQDKLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWSKNGLESVVK-----RDPYGKSSRIRA 755
QY 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVTFQIGSVF 813
Db 756 STGVGFQWQSPIGPLFVSYAKPIKKYENDDDVDFQFQSIGSF 797
```

RESULT 2

```
US-08-433-522A-4
; Sequence 4, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
```

```
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-4
```

```
Query Match 25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY 23 STHAQAADFMANDTITGLQVRVTIESLOSVPFRGLGVVSENQADGVKALYATGNFSDV 82
Db 14 TTTVFAAPFAKDIRVDGVGDLEQOIRASLPVRACORVTDNDVANIVRSLSVSGRFDV 73
QY 83 QVYHOGRI-IYQVTERPLIAEFNFGNRLIPKEGLQGLKNAGLAVGQPLKQATVOMIE 141
Db 74 KA-HQEGDLVWSVAKSIISDVKIKGNSVITPEALKQNDLNGKFKYGVLDLIREKNEFA 132
QY 142 TELTNQYISQGYNTTEITVQOTMLDGNRVKLDMTFAEGKPARVVDINIIGNQHSFADLI 201
Db 133 KSVKHYASVGRYNATPEIVNTLPNNRAELIQINEDDRKAKLASLTFKGNESVSSTLQ 192
QY 202 DVLAIKDNKINPLSKADRYTQELVTSLENLRAKYNAGVFRFEIKDAKLINEDEKKNRIF 261
Db 193 EQMELQPSWKKL-WGNKFEQAQFKDQSIIRDYLYNNGYAKAQITKTDVQNDKTKVN 251
QY 262 VEISLHEGEQYRFQOTFLGNLTQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLOYSARIIIGNLGMSAELEPLLSALHLNDFRSDIADVENAIKALGE 311
QY 321 DGYYAQRVTRINDESRTVDVEYIDPVHPVYVRINFTGNFKTQDEVLRREMRLQEG 380
Db 312 RYGSATVNSVPDFFDANKLTAITLVVDAGRRLTVRQLRFGNTVSADSLRQEMRQEG 371
QY 381 ALASNQIKLSRARLMRTGFKHVTVDTR--PVPNSPDQDVNVFVVEQPSGSSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTPE--TVENRIDPINGSNDEVDVYKVKERNITGSINFGIGY 429
QY 439 SOSGGVTFQFDVSONNFMGTGKHVNASFRRSETREYVSLGMTNPNYTVNGVCSLSGYR 498
Db 430 GTESGISYQASVKQDNFLGTGAASIACTKNDYGTSVNLGYTEPYFTKDG--SLGGNVF 487
QY 499 KTKYDNK---NISNVLDSDYSGSLSYGYPIDENORISFGL-----NADNTKLHGGRFMI 550
Db 488 FENDNKSDDTSNRYKRTYGSNLTGFPVNNNSYVGLGHTYKNKISNPALEYNRNLYI 547
QY 551 SNVKQLMADGKGIQVNDNGIPDFKHDTYTNAILGNWYSSLDPRVFPQOGMSHVD--LT 608
Db 548 QSMK-----FKGNGIKTNDDFS-----FGWYNSLNRGYFTPKGVKASLGGGRVT 592
QY 609 VGFGDKTHOKVYVQGNIRPE-----IKKSVLRGYAKLGYGN-NLPFFENFVAGGYGS 660
```


Db 593 IPGSDNRYKLSADVOGFYPLDRDLHWVWSAKASAGYAN-GFGNKRPLFPYQTYTAGGIGS 651
Qy 661 VRGYDQSSLPGRSQAYLTARRGQQTTLG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGSGIPNA---IYAEYNGSGTGTFKKISSDVIIGGNATATASAEELIVPTFFVSDK 708
Qy 714 I-DQVRPVIFIEGGQVFDTT-GMDKQTDLTQFKDPQATAEQNAKAAANRPLLTQDKQLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYGKSSRIRA.755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVDEQFQFSGGSF 797

RESULT 3

US-08-433-522A-6
; Sequence 6, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-433-522A-6

Query Match 25.08; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

Qy 23 STHQAADFMANDITITGLQRVTIESLQSVLPFRIGVQVSENQADGVKALYATGNFSDV 82
Db 14 TTTVFAAPFVAKDIRVDSQVGLDLEQQIRASLPVRGQRTVDNVANIVRSLEFVSGREFDV 73
Qy 83 QVYHOGRI-IYQYTERPLAIEFEGNRLIPKEGLQBLNAGLAVQPLKQATVQOMIE 141
Db 74 KA-HQEGDVLVSVVAKSIISDVKIGNSVIPTBALKONLDANGFKVGDVIREKLNEFA 132

Qy 142 TELTNOYISQYINTEITVVKOTMLDGNRVKLDMTFAEGKPARVVDVINIIGNQHFSADADI 201
Db 133 KSVKEHYASVGRNATVEPIVNTLPNNRAEILQINEDDKAKLASLTFKGNESVSSSTLQ 192
Qy 202 DVLAKDNKINPLSKADRYTQEKLVTSLENRAKYLNGAFVRFKDKAKLTNEDKNRIF 261
Db 193 BOMELQDPSMWKL-WGNKFEQAQFEKLOSIDRYLLNNGYAKAQITKTQVLDNDEKTKVN 251
Qy 262 VEISLHGEQYRFGQTQFLGNTTYTQAELEALLK-FKAEEGFSOAMLEOTTNNISTKFGD 320
Db 252 VTIDVNEGLQDLRSARIIGNLGMASLEPLLSALHLNDTFRSDIADVEMAIKALGE 311
Qy 321 DGYTAAQIRPVTRINDESRITVDVEYIIDPVHPVYVRRINFTGNFKTQDEVLRMRQLEG 380
Db 312 RGYGSATVNSVPDEDDANKTLAITLVVDAGRRLTVRQLRFEQNTVSADSTLRQENRQEG 371
Qy 381 ALASNQIKQLSRALMRTPGFFKHVTVDR--PVPNSPOQVDVNVFVVEQSPSSSIAAGY 438
Db 372 TWYNSQLVELGKIRLDRTGFFE--TVENRIDPINGSNDEVDVYVKVKERTNGTSINFGIGY 429
Qy 439 SOSGGVTFQFDVSONNFMGTGKHVNASFSSRSETRVYSLGMTNPYFTVNGVSQSLSGYR 498
Db 430 GTESGISYQASVKQDNFLCTGAANSIAGTKNDYGTSVNLGYTEPYFTKDG--SLGGNVF 487
Qy 499 KTKYDNK---NISNYVLDISYGSLSYGYPIDENQRISFCL-----NADNTKLHGGREFMI 550
Db 488 FENYDNSKSDTSSNYKRTYTGSNVTGLGPPVNNNSYVVLGHTYNNKISNFALEYNRNLYI 547
Qy 551 SNVQLMADGGKIQVDNNGIPDFKHDTYTYNAILGNWYSSLDPRVPFPTQGNHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FCWNTNSLNRGYEPTKGVKASLGGRVT 592
Qy 609 VFGDKTHQKVYQGNIVRPF-----IKKSVLRGYAKLGYGN-NLPYENFYAGGYGS 660
Db 593 IPGSDNRYKLSADVOGFYPLDRDLHWVWSAKASAGYAN-GFGNKRPLFPYQTYTAGGIGS 651
Qy 661 VRGYDQSSLPGRSQAYLTARRGQQTTLG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGSGIPNA---IYAEYNGSGTGTFKKISSDVIIGGNATATASAEELIVPTFFVSDK 708
Qy 714 I-DQVRPVIFIEGGQVFDTT-GMDKQTDLTQFKDPQATAEQNAKAAANRPLLTQDKQLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLSDVLK-----RLPDYGKSSRIRA 755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVDEQFQFSGGSF 797

RESULT 4

US-09-135-166-2
; Sequence 2, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0; Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-135-166-2

Query Match      25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY 23 STHAAQADFMANDITITGLQRTVIESLQSVLPFRLGQVSVENQADGVKALYATGNFSDV 82
DB 14 TTTVFAAPFVAKDIRVDGVQGDLEQOIRASLPVRAGORTDNDVANIVRSFLVSGRFDV 73
QY 83 QYVHQEGRI-IYQVTERPLIAEINFEGNRLIPKEGLEGLKNAGLAVGQPLKQATQMIE 141
DB 74 KA-HQEGDLVSVVAKSIISDVKIKGNSVPTPEALKNLDANGFKVGVDLIREKLENEFA 132
QY 142 TELNQYISQGYNTTEITVKTMDGNRYKLDMTFAEKPARYVDVINIGNHFSADALI 201
DB 133 KSVREHYASVRYNATPEIVNTLPNNRAETILQINEDDKAKLASLTFKGNESVSSSTLQ 192
QY 202 DVLAIKONKINPLSKADRYTQEKLVTSLENLRKALYAGFVRFEKDAKLINEDKNRIF 261
DB 193 EQMELQDPSWKL-WGNKFEAGQFEKLOSIIRDYLNNGYAKAQITKDVQINDKTKVN 251
QY 262 VEISLHEGEQYRFQOTQFLGNLTQVQAEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
DB 252 VTIDVNEGLQDLSARLIIGNLGNLSAELEPLLSALHLNDTFRSDIADVENAIKAKLGE 311
QY 321 DGYTAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFKTQDEVLRREMRQLEG 380
DB 312 RYGSATVNSVPDFDDANKTLAITLVDAGRRLTVRQLRFEQNTVSADSTLRQEMRQOEG 371
QY 381 ALASNOKTLQSLARLMRTGFFKHVTVDTR--PVPNSPDQVDVNFVVEQPSGSSSTIAAGY 438
DB 372 TWYNSQLVELGKIRLDRTGFFE--IVENRIDPINGSDEVDVYKVKERNGTSGINFGIGY 429
QY 439 SOSGVTTFQFVSQNFQMGTKHVNASFSETRRETVYSLGTMNPYFTVNGYVSQSLSGYR 498
DB 430 GTEISISYQASVKQDNFLGTGAAGVSIAGTKNDYGTGVNLGYTEPYFTKDGV--SLGGNVF 487
QY 499 KTKYDNK---NISNVVLSYSGSLSGYPIDENQIRISFGL-----NADNTKLHGGRFNGI 550
DB 488 FENYDNKSDTSNRYKRTYGSNVTLGFPVNNENSYVGLGHTYKNISNFALEYNRNLVI 547
QY 551 SNVKQLMADGKIQVDNNGIPDKHDYTYNAILGNWTSLSLDRPVFPQTQMSHSVD--LT 608
DB 548 QSMK-----FKGNGIKTNDKFS-----FGWYNSLNRGYFPFKGVKASLGRRVT 592
QY 609 VGFQDKTHQVYVYQGNINRPF-----IKKSVLRGYAKLGYGN-NLPYENFYAGGYGS 660
DB 593 IPGSDNKYKLSADVQGFYPLDRDHLWVVSASAKASAGYAN-GFGNKRPLFPYQTYTAGGIGS 651
```

```

QY 661 VRGYDQSSLSGRSQAYLTARRGQQTTLG-----EYVGNALATFSGSELILPLPFKGDW 713
DB 652 LRGFAYGSGIGNA--IYAEVGNSSGTGTFKKISSDVIIGNAIATASAEILVPTPFVSDK 708
QY 714 I-DQVRPVIFIEGGQVDTT--GMDKQITDILTQFKDPQATAEQNAKAANRPLLTQDKQLRY 771
DB 709 SONTVRTSLFVDAASVWNTKWSKNGLESVLR-----RLPDYGKSSRIRA 755
QY 772 SAGVGATWYTPIGPLSISYAKPLNKKONDOTDTVQFOIGSVF 813
DB 756 STGVGFQWQSPGIGPLVFSYAKPIKKYENDVDVEQFQFSIGGSF 797

RESULT 5
US-09-135-166-4
; Sequence 4, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-135-166-4

Query Match      25.0%; Score 1052.5; DB 3; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;
```



```
QY 661 VRGYDQSSSLGRSQAYLTARRGQTTTG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGISGPN-----IYAEGNGSGTGTFFKKISSDVIGGNATATASAEILVPTPFVSDK 708
QY 714 I-DQVRPVIFIEGGQVFDTT-GMDKQTDLTQFKDPQATAEQNAKAAANRPLLTQDKOLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWSKNGLESVDLK-----RLPDYKSSSRIRA 755
QY 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQDQDTVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVDEQFQFSGGSF 797

RESULT 7
US-08-942-046-2
; Sequence 2, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942.046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 797 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-942-046-2
Query Match 25.0%; Score 1052.5; DB 4; Length 797;
Best Local Similarity 31.6%; Pred. No. 4.5e-76;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;
QY 23 STHAQADFMANDTITIGLQVRVTIESLQSLVPLFRLCQVVSQNLADGVKALYATGNFSDV 82
Db 14 TTTTFAAPFAVKDIRVDGVQGLEQIIRASLPVRAGQRTDNDVANIVRSFVSGRFDV 73
QY 83 QYTHQEGRI-IYQVTERPLIAINEFGRNLIPKEGLQGLKKNAGLAVGQPLKQATVQMIE 141
```

```
Db 74 KA-HQEGDVLVWSVAKSIISDVKIKGNSVIPTKALKNLDANGFKVGDVLIRKLENEFA 132
QY 142 TELTNOYISOGYNTTEITVKTMLDGNRVKLDMTFAGKPARVVDIINIGHQHESDADLI 201
Db 133 KSVKEHVASGRYNATVEPIVNTLPNRAEILIOINEDDRAKAKLASLTFKGNESVSSTLQ 192
QY 202 DVLAIDKNKINPLSKADRYTQEKLVTSLENLRAKYLNAGVFVFETIKDAKLNINEDKNRIF 261
Db 193 EQMELQDPSWWKL-WGNKFEQAQFEKDLQSRDYLYLNGYAKAQITKTQDVLNDEKTKVN 251
QY 262 VEISLHEGEQYRFQTOFLNLTYYQAELEALLK-FRAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRSDIADVENAKIKLIGE 311
QY 321 DGYYYAQRVPVTRINDESRTVDVEYIDPVHPVVRVRRINFTGNFKTQDEVLRRMRQLEG 380
Db 312 RYGSATVNSVPDDEDDANKILAILVDVAGRRLIVRQLRFEGNTVSADSTLRQEMRQEG 371
QY 381 ALASNQIKQLSRARLMRTGFFKHVTVDT--PVPNSPDQVDVNFVVEPQPSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGPE--TVENRIDPINGSNDEVDVYKVKERNGTGSINFGIGY 429
QY 439 SOSGVTFQFDVYSONNPMGTGKHVNASFSRSETREYVSLGTMNXPYVTNGYSQSLSGYR 498
Db 430 GTESGISYQASQKQDNFLGTGAASVIACTKNDYGTSMNLGTYTEPFTTKDGV--SLGGNVF 487
QY 499 KTKYDNK---NISNVLDISYSGSLSYGYPIDENORISFGL-----NADNTKLHGRFMGI 550
Db 488 FENYDNKSDTSSNYKRTYGSNTVLCPPVNNNSYVVLGHTYNTKISFALEYNRNLYI 547
QY 551 SNVKQLMADGGKIQVDNNGIIPDFKHDTTYNAILGWNYSLSLDRPVPFQOGMSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDFDES-----FGWNTNSLNRGYFTPKGVKASLGGGRVT 592
QY 609 VFGDKTHQKVVYQGNLYRPF-----IKKSVLRGYAKLGYGN-NLPFYENFYAGGYGS 660
Db 593 IPGSDNKYIKLSADVOGFYPLDRDLHWVSASAKAGYAN-GFGNKKLFFYQTYTAGGGS 651
QY 661 VRGYDQSSSLGRSQAYLTARRGQTTTG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGISGPN-----IYAEGNGSGTGTFFKKISSDVIGGNATATASAEILVPTPFVSDK 708
QY 714 I-DQVRPVIFIEGGQVFDTT-GMDKQTDLTQFKDPQATAEQNAKAAANRPLLTQDKOLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWSKNGLESVDLK-----RLPDYKSSSRIRA 755
QY 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQDQDTVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVDEQFQFSGGSF 797
```

```
RESULT 8
US-08-942-046-4
; Sequence 4, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```


Db 74 KA-HQEGDVLVSVAKSIISDVKIKGNSVIPTEALKQNDLNGFVGVGLIREKLNEFA 132
QY 142 TELTNOYISOGYVNTTEITVKQTMLDGNRVKLDMTFAGKPARVVDINTIGNQHFSDADLI 201
Db 133 KSVKEHYASVGRYATVPEIVNTLPNNRAILIQINEDDKAKLASLTFKGNESVSSSTLQ 192
QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNLRKAKYINAGVREIKDAKININEDKNRIF 261
Db 193 EQMELQPDSSMWKL-WGNKFEAGFQKDLQSIIRYIYLNNGYAKAIIKTVDVQNDKTKVN 251
QY 262 VEISLHEGEQYRFGQTOFLGNLTQAELEALLK-FKAEFGFSQAMLEQTTNNISKFGD 320
Db 252 VTIDVNEGLOYDLRSARIIGNLGSMAELPILLSALHLNDFRSDIADVENAIKALGE 311
QY 321 DGYVYQAIQIRPVTRINDESRVDEYIDPVHPVYVVRINFTGNFKTQDEVLRREMRQLEG 380
Db 312 RYGSATVNSVPDFFDANKTLAITVVDAGRRLLTVRQLRFEAGTVSADSTLRQEMRQEG 371
QY 381 ALASNOKIOLSRARLMRTGFFKHVTVDTR--PVPNSPDQVDVNFVVEEQPSGSSSTIAAGY 438
Db 372 TWNSQLVELCKIRLDRDTGFFE--TVENRIDPINGSNDVYVVKYKERTGTSINFGIGY 429
QY 439 SOSGGVTFQFDVSONNPMGTGKHVNASFSRSRETVYSLGWTNPYFTVNGVSQSLSGYR 498
Db 430 GTEGSIYSTQSIKQDNFLGTGAASVIAGTNDYCTSVNLGYTEPYFTKDGV--SLGGNVF 487
QY 499 KTKYDNK---NISNYVLDSSGGSLSGYGPIDENQRIISFGL-----NADNTKLHGGRMGI 550
Db 488 FENYDNKSDTSSTSNKYKRTTYSNVTLGFPVNNNSYVGLGHYKNSNFALEYNRNLYI 547
QY 551 SNVKQLMADGGKIQVNDNGIPDFKHDYTTNAILGNWYSSLDPRVFPPTQGMHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYFPTKGVKASLGGRVT 592
QY 609 VGFGDKTHQKVYV--YQG-----NIRPFTKKSVLGAKLGYCN-NLPFENFYAGGYGS 660
Db 593 IPGSDNKYYKLSADVQGYPLDRDLHVMVSAKASAGYAN-GFGNKRLPFYOTYTAGGIGS 651
QY 661 VRGYDQSSGLGRSQAYLTARRGQOTTIG-----EVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGPAYGISGPN---IYAEVNGSGTGTFKKISSDVIGGNATASAEILVPTPFVSDK 708
QY 714 I-DQVRPVIFTEGQVFDTT--GMDKQITDLTQFKDQATBQNAKAAANRPLLTQDKQLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWSKNGLESVLC-----RLPDYGRKSSIRA 755
QY 772 SAGVATWYTPIGPLSISYAKPLNKKQNDQTDVTQFOIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDVVEQFOFSIGGSF 797

RESULT 10
US-08-433-522A-10
; Sequence 10, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-Sep-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-10

Query Match 24.9%; Score 1047.5; DB 3; Length 793;
Best Local Similarity 31.7%; Pred. No. 1.1e-75;
Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;

QY 23 STHAAADAFMNDITITGLQRTVIESLQSVLPFRFGVSVSENQADGVKALYATGNFSDV 82
Db 14 TTVFAAFVFKDLRVDGVQGDLEQOIRASLPVRAGORVTDNDVNIYVRSFLVSGREDDV 73
QY 83 QVYHOEGRI-IYQVTERPLIAEINFEGRNRLIPKEGLQBLKAGLAVGOPLKQATVQMI 141
Db 74 KA-HQEGDVLVSVVAKSIISDVKIKGNSVIPTEALKQNDLNGFVGVGLIREKLNEFA 132
QY 142 TELTNOYISOGYVNTTEITVKQTMLDGNRVKLDMTFAGKPARVVDINTIGNQHFSDADLI 201
Db 133 KSVKEHYASVGRYATVPEIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSSTLQ 192
QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNLRKAKYINAGVREIKDAKININEDKNRIF 261
Db 193 EQMELQPDSSMWKL-WGNKFEAGFQKDLQAIIRDYVNLNNGYAKAIIKTVDVQNDKTKVN 251
QY 262 VEISLHEGEQYRFGQTOFLGNLTQAELEALLK-FKAEFGFSQAMLEQTTNNISKFGD 320
Db 252 VTIDVNEGLOYDLRSARIIGNLGSMAELPILLSALHLNDFRSDIADVENAIKALGE 311
QY 321 DGYVYQAIQIRPVTRINDESRVDEYIDPVHPVYVVRINFTGNFKTQDEVLRREMRQLEG 380
Db 312 RYGNVTVNSVPDFFDANKTLAITVVDAGRRLLTVRQLRFEAGTVSADSTLRQEMRQEG 371
QY 381 ALASNOKIOLSRARLMRTGFFKHVTVDTR--PVPNSPDQVDVNFVVEEQPSGSSSTIAAGY 438
Db 372 TWNSQLVELCKIRLDRDTGFFE--TVENRIDPINGSNDVYVVKYKERTGTSINFGIGY 429
QY 439 SOSGGVTFQFDVSONNPMGTGKHVNASFSRSRETVYSLGWTNPYFTVNGVSQSLSGYR 498
Db 430 GTEGSIYSTQSIKQDNFLGTGAASVIAGTNDYCTSVNLGYTEPYFTKDGV--SLGGNVF 487
QY 499 KTKYDNK---NISNYVLDSSGGSLSGYGPIDENQRIISFGL-----NADNTKLHGGRMGI 550
Db 488 FENYDNKSDTSSTSNKYKRTTYSNVTLGFPVNNNSYVGLGHYKNSNFALEYNRNLYI 547
QY 551 SNVKQLMADGGKIQVNDNGIPDFKHDYTTNAILGNWYSSLDPRVFPPTQGMHSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDDFS-----FGWYNSLNRGYFPTKGVKASLGGRVT 592
QY 609 VGFGDKTHQKVYV--YQG-----NIRPFTKKSVLGAKLGYCN-NLPFENFYAGGYGS 660
Db 593 IPGSDNKYYKLSADVQGYPLDRDLHVMVSAKASAGYAN-GFGNKRLPFYOTYTAGGIGS 651
QY 661 VRGYDQSSGLGRSQAYLTARRGQOTTIG-----EVVGGNALATFGSELILPLPFKGDW 713

Db 652 LRGFAYGSGPNA---IYAEHNGTFFNKISSDVIGGNAITTAASAEILVPTFVSDKQNT 708
Qy 717 VRPVIFIEGGVDFDT-GMDKQTDLTQFKDQPAQAEQNAKAAANRPLLTQDKOLRYSGV 775
Db 709 VRTSLFVDAASVWNTKSKNGLESVKLD-----LPDYGKSSRIRASTGV 755
Qy 776 GATWYTPIGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSGSF 793
RESULT 11
US-09-135-166-10
; Sequence 10, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-135-166-10

Query Match 24.98; Score 1047.5; DB 3; Length 793;
Best Local Similarity 31.78; Pred. No. 1.1e-75;
Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;
Qy 23 STHQAADFMANDTITGLQRTVTTESLQSVLPFRGLGVVSENQLADGVKALYATGNFSDV 82
Db 14 TTTVFAAPFVFPKRDVGVQGLQEQIRASLFPVRAGQRTDNDVANIYRSLFVSGRFDV 73
Qy 83 QVYHQEGRI-IVQVTERPLIAEINFEGNRLIPKPEGLQSLKAGLAVGQPLKQATVQME 141
Db 74 KA-HQEGDVLVVSVVAKSIIDSVKIGNSVIPTEALKQNDLQANGFKVGDVLIREKLNFEA 132

Qy 142 TELTNOYISQYNTFETVTKQTMLDGNRVKLDMTAEAGKPARVVDINIIGNQHFSADALI 201
Db 133 KSVKEHYASVGRV NATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSTLQ 192
Qy 202 DVLATKDKINPLSKADRYTQEKLVTSLENRAKYLNAGFVRFEEKADKLINEDKNRIF 261
Db 193 EOMELQPSDWMKL-WGNKFEQAQFEKDLQAIRDYLLNNGYAKAQITKTQDVLDNDEKTKVN 251
Qy 262 VEISLHEGEQYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDNGLQYDLRSARIIGNLGGMSAELEPLLSALHLNDTFRSDIADVENAIKAKLGE 311
Qy 321 DGYVYAIQRPVTRINDESRVDVEYYIDPVHVVYVRRINFTGNFTQDQVLRREMRQLG 380
Db 312.RGYGNTTVNSVPDFDDANKTLAITFVVDAGRRLTVRQLRFEGNTVSADSTLQENRQSG 371
Qy 381 ALASNQKIQLSRARLMRTGFFKHVTVDR--PVPNSPOVDVNVFVVEQPSGSSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYVVKVKTERTGSINFGIGY 429
Qy 439 SQSGGVTFQFDVSONNFMGTGKHNASFSRSETREVYSIGMTNPYFTVNGVSQSLSGYYR 498
Db 430 GTESGISYQTSIKQDNFLCTGAANSIAGTKNDYGTSVNLGYTEPYFTKDGV--SLGNIF 487
Qy 499 KTYDNK---NISNYVLDYSGLSYGPIDENQRISFGL-----NADNTKLHGGRFMGI 550
Db 488 FENYDNSKSDTSSNYKRTYGSNVTLPFPVNNNSYVVLGHTYKNKISNFALEYNNLYI 547
Qy 551 SNVKOLMADGGKIQVDNNGIPDFKHDYTYNNAILGNWYSSSLDRPVFPQGMHSVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWNTNSLNRGYEPTKGVKASLGGRYT 592
Qy 609 VGFGDKTHQVY--YOG-----NIYRPTFKSVLRGYAKLGYGN-NLDPYENFYAGGYGS 660
Db 593 IPGSDNKYKLSADVQGFYPLDRDRHRWVWSAKASAGYAN-GGKNRLFPYQYTTAGGTGS 651
Qy 661 VRGYDQSSGLSPRSQAYLTARRQQT---TLGEVVGNALATFGSELILPLPFKGDWI-DQ 716
Db 652 LRGFAYGSGPNA---IYAEHNGTFFNKISSDVIGGNAITTAASAEILVPTFVSDKQNT 708
Qy 717 VRPVIFIEGGVDFDT-GMDKQTDLTQFKDQPAQAEQNAKAAANRPLLTQDKOLRYSGV 775
Db 709 VRTSLFVDAASVWNTKSKNGLESVKLD-----LPDYGKSSRIRASTGV 755
Qy 776 GATWYTPIGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDDDVEQFQFSGSF 793
RESULT 12
US-08-942-046-10
; Sequence 10, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 793 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-942-046-10

Query Match 24.9%; Score 1047.5; DB 4; Length 793;
Best Local Similarity 31.7%; Pred. No. 1.le-75;
Matches 259; Conservative 160; Mismatches 334; Indels 65; Gaps 21;

QY 23 STHQAADFMANDITITGLQVTTIESLQSVLPFRGLQVVSSENOLADGVKALYATGNFSDV 82
Db 14 TTVFAAPFPVKIRVDVGQGLEQIRASLPVRAGQRTDNDVANIVRSFLVSGRFDV 73

QY 83 QYVHOEGRI-IYQVTERPLIAEINFEGRNLIPKEGLQGLKNAGLAVQPLKQATVOMIE 141
Db 74 KA-HQESDVLVSVAKSIISDVKIKGNSIIPPEALKQNLNDANGFKVGDVILIREKLENEFA 132

QY 142 TELTNOYISQGYNTYETITVKQTMIDGNRKLDMTFAEKGPARVVDINIGNHQFSDADLI 201
Db 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLITFKGNSVSSSTLQ 192

QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLLENRAKYLNAGFVRFPEIKDAKLINEDKNRIF 261
Db 193 EQMELQPDSSWWKL-WGNKFEQAQFEKDLQAIRDYLLNNGYAKAQITKTVDVNLDEKTKVN 251

QY 262 VEISLHEGEOYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTNNISTKFKGD 320
Db 252 VTIDWEGQLYDLRSARIIIGNLGGMSAELEPLLSALHLNDFRSDIADVENAIKALGE 311

QY 321 DGYTAAQIRPVTRINDESRVDVEYIIDFVHPVYVRRINFTGNFKTQDEVLRREMRQLEG 380
Db 312 RYGNNTVNSVPDEDDANKTLAITFVDAGRLTIVRQLRFEQNTVSADSTLRQEMRQOEG 371

QY 381 ALASNOKTOLSRARLMRTGFFKHVTVDIR--VPVNSPOVDVNFVVEEQPSGSSTIAAGY 438
Db 372 TWNSQLVELGRIDRTGTFE--TVENRIDPINGSNDEVDVYVVKERTGNTSIFNGIGY 429

QY 439 SOSGGVTFQFQYSONFMGTGKHVNASFSRSETREYISLGMTNIPYFTVNGVSQSLSGYYR 498
Db 430 GTEGSIYQTSYTKQDNFLGTGAVSAGTAKNDYGTISVNLGYETPYTKDGV--SLGGNIF 487

QY 499 KTKYDNK---NYSNVLDSYSGSLSYGYPIDENQRIISFGL-----NADNTKLUGGRFMI 550
Db 488 FENYDNKSQDTSNRYKRTYGSNVTGLGFPNNNSYVGLGHTYNNKISNFALEYNRNLXY 547

QY 551 SNVKQLMADGGKIQVDNNGIPDKHDTYTNALTLGNYSLLDRPVPPTQCMGSHSDV--LT 608
Db 548 QSMK-----FKNGIKTNDFFDS;----FGWNYSNLNRCGYPTTKGVKASLGRVT 592

QY 609 VGFQDXTOKV--YQG-----NIYRPFKIKSVLRGAKLGYGN-NLPFYENFYAGGYS 660
Db 593 IPGSDNKKYKLSADVQGFPLDRDRHWVWSAKASAGYAN-GFGNKRLLPFQVTVTAGGIGS 651
```

```

QY 661 VRGYDQSSGLPSRQAYLTARRGOQT-----TIGEVVVGGNALATFGSELILPLPFKGDWI-DQ 716
Db 652 LRGFAYGSIQPN---IYAEHNGTGNKISSDVTGNGNAITTSAAELIVPTFPVSDKSQNT 708

QY 717 VRPVIFIEGQVFDTT-GMDKQTFIDLTQFKDPQATQNAQNAKAAANRPLLTQDKQLRYSGV 775
Db 709 VRTSLFVDAASVWNTKWKSDKNGLSKVLKD-----LPDYGKSSRIASTGV 755

QY 776 GATWYTPIGPLSISYAKPLNKKQNDQTDVTQVQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDNDVQFQFSIGGSF 793

RESULT 13
US-08-433-522A-8
; Sequence 8, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 792 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-433-522A-8

Query Match 24.8%; Score 1042; DB 3; Length 792;
Best Local Similarity 31.4%; Pred. No. 3.le-75;
Matches 257; Conservative 159; Mismatches 335; Indels 68; Gaps 20;

QY 23 STHQAADFMANDITITGLQVTTIESLQSVLPFRGLQVVSSENOLADGVKALYATGNFSDV 82
Db 14 TTVFAAPFPVKIRVDVGQGLEQIRASLPVRAGQRTDNDVANIVRSFLVSGRFDV 73

QY 83 QYVHOEGRI-IYQVTERPLIAEINFEGRNLIPKEGLQGLKNAGLAVQPLKQATVOMIE 141
Db 74 KA-HQESDVLVSVAKSIISDVKIKGNSIIPPEALKQNLNDANGFKVGDVILIREKLENEFA 132

QY 142 TELTNOYISQGYNTYETITVKQTMIDGNRKLDMTFAEKGPARVVDINIGNHQFSDADLI 201
Db 133 QSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLITFKGNSVSSSTLQ 192
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:51:21 ; Search time 69.95 Seconds
(without alignments)
1116.807 Million cell updates/sec

Title: US-09-701-711-2
Perfect score: 4202
Sequence: 1 MRSYFKGFQVSAMTMAVM.....LNKKQNDQDTVQFQIGSVF 813
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1451.5	34.5	797	2 H83190	probable outer mem
2	1262	30.0	795	2 AC0129	probable surface a
3	1218.5	29.0	803	2 AB0530	outer membrane pro
4	1210	28.8	810	2 A64742	hypothetical prote
5	1210	28.8	810	2 C90651	hypothetical prote
6	1210	28.8	810	2 C85052	hypothetical prote
7	1203	28.6	803	2 B82099	surface antigen VC
8	1172.5	27.9	784	2 E82731	outer membrane ant
9	1115	26.5	797	2 C81228	outer membrane pro
10	1111	26.4	797	2 D82000	outer membrane pro
11	1053.5	25.1	808	2 F64102	protective surface
12	1052.5	25.0	797	2 JC4078	protective surface
13	754	17.9	768	2 D71726	outer membrane pro
14	734	17.5	774	2 D97527	omp1 protein precu
15	734	17.5	774	2 AE2746	group 1 outer memb
16	727.5	17.3	769	2 F87486	outer membrane pro
17	697	16.6	768	2 B97725	outer membrane pro
18	680.5	16.2	781	2 AH3355	outer membrane pro
19	641	15.3	617	2 H84957	hypothetical prote
20	568.5	13.5	778	2 C70412	outer membrane pro
21	515	12.3	739	2 A81430	outer membrane pro
22	434	10.3	916	2 G64601	protective surface
23	429	10.2	906	2 F71910	probable outer mem
24	397.5	9.5	792	2 H81693	outer membrane pro
25	396.5	9.4	792	2 B71539	probable omp85 ana
26	395	9.4	790	2 D86528	omp85 analog [impo
27	395	9.4	790	2 D72094	omp85 analog - Chl
28	336.5	8.5	853	2 A71339	probable outer mem
29	325	7.7	821	2 B70199	outer membrane pro

ALIGNMENTS

RESULT 1

H83190

Probable outer membrane protein PA3648 [imported] - Pseudomonas aeruginosa (strain P
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83190
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
Nature 406, 959-964, 2000
A:Reference number: AB2950; MUID:20437337
A:Accession: H83190
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-797 <STO>
A:Cross-references: GB:AB004784; GB:AE004091; NID:g9949799; PIDN:AAG07036.1; GSPDB:G
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3648
C:Superfamily: protective surface antigen D-15

30	271	6.4	833	2	AF2089	hypothetical prote
31	256.5	6.1	578	2	B64012	hypothetical prote
32	256	6.1	853	2	D70304	hypothetical prote
33	254.5	6.1	617	2	AF2897	conserved hypot het
34	254.5	6.1	641	2	H97672	hypothetical prote
35	251.5	6.0	676	2	AE2417	hypothetical prote
36	249	5.9	711	2	A72375	hypothetical prote
37	236.5	5.6	846	2	F75525	outer membrane pro
38	230.5	5.5	578	2	AE0428	probable exported
39	230	5.5	577	2	AD1058	probable exported
40	228	5.4	577	2	F86119	hypothetical prote
41	227	5.4	577	2	S56445	hypothetical 64.8K
42	227	5.4	577	2	F91278	hypothetical prote
43	224.5	5.3	861	2	S77409	hypothetical prote
44	221.5	5.3	579	2	F83327	conserved hypot het
45	198	4.7	623	2	AI3488	outer membrane pro

Db 359 TEDEVLRREMRQEGGWASLYILIDQSKARLERLGKFEVNEVPANPGTDQDVVNSVE 418
QY 426 EQPSGSSTIAAGYSQSGVTFQFDVSONPMGTCKKHNASFSRSETRVYSLGNTDNYFT 485
Db 419 EQPSGSITASVGAQSAAGLIGGSIQNNELGTGKVKVSIIGLIRSEYQTRYNFGVDPIYT 478
QY 486 VNGVQSGLSYYRKTVD--NKNISNVYLDVSYGSSLSYGYPIDENQRIISGLNADNKLH 543
Db 479 VDGVSGLYNAFYKRTDYDELVDVSYVNSVSLGAGMSIGYPISTSLTTLGLSVQRDQID 538
QY 544 GGFMGTSNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSILDRPVFTQGMH 603
Db 539 TGRIT-VDEIYDFLDKEG-----DN-----FTNEKASIGWSESLNKGVLATRGHSQ 584
QY 604 SVDL--TVGFGDKTHQKVVYOGNIYRPIKKSVLRGYAKLGYGN-----NLPIYENFYA 655
Db 585 SLTLETTLPGLSDFYKIDYRGQVAFPLTDNYTMRFTLGYGDSYSTERLPYENFYA 644
QY 656 GGYGSRVGYDSSGLGPRSOAVLTARRGQQTTL-----GEVVGGNALAT 698
Db 645 GGFNSVRGFKDSTLGRSTPSV-ARNPDGTPKKNQGPSKGRYTDPODPAFGNLTIT 703
QY 699 FGSLLILPLPFKGDWIDQVREVIIEGQVEDTTGMDKQTDLTQFKDPQATASQNAKA 758
Db 704 GGAELLPLPFVKDQ-ROLRTVLFWDVGSTEDTDCPTKTTTNCDBIK-----749
QY 759 NRPLLTQDKQLRYSGAGCATWYTPIGPLSISYAKPLNKKONDQDTVQFOIGSVF 813
Db 750 -----TDNLASSYGVGLTWITALGPLSLATPIKPDNAETQVQFSLGQTF 797

RESULT 2
AC0129
probable surface antigen YP01052 [imported] - Yersinia pestis (strain C092)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AC0129
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0129
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-795 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC89894.1; PID:gl59791119; GSPDB:GN00175
C:Gene: YP01052
C:Superfamily: protective surface antigen D-15

Query Match 30.0%; Score 1262; DB 2; Length 795;
Best Local Similarity 35.7%; Pred. No. 3.8e-65;
Matches 292; Conservative 152; Mismatches 329; Indels 44; Gaps 14;
QY 14 MTMAVMMVMS-----THAQADFMANDITITGLQRTVIESLQSVLPFRGQVYSENOLADGVKA 72
Db 6 LIASLLFSGSVYAGDGFVNDIHFEGQLQVAVGAALLNMPVRVGDVTSDDDIGKTIRA 65
QY 73 LYATGNFSDVQVYHQEGRIYQVTERPLIAEINFEGRNRLIPKEGLEGLKNAGLAVGQPL 132
Db 66 LEATGNFEDVRVLDGNTLIVQVKERTIASITFSGNKAVKEDMDLKNLEASGVRVGEAL 125
QY 133 KQATVOMIETELNOVISQYNTETIVKOTMLDGNRKLDMTFAEGKPARVVDINIIGN 192
Db 126 DRTTISNIEKLEGLD-----GVKYSKASVAVVTPLPNRVLDKIVFTECVSAKIQIINIVGN 185
QY 193 QHPSADLIDVLAIMD-----NKNIPNLSKADR-YTOEKLVTSLENURKAYLNAGFVRFEI 246
Db 186 HSTTDELIISRFQLRDVRPWNVV-----GDRKYQKOKLAGDLETLRSLFRYLDGRYARFNI 240

QY 247 KDAKLINEDKNRIFVEISLHEGEQYRFQOTFLGNITTYTOAELEALLKKAEBGFSQAM 306
Db 241 DSTQVSLTPDKKGIYVITINITEGPOFKLSNVISGNLAGHQSEAKLTIPEGLFNGSK 300
QY 307 LEQTTNISTKFGDDGYIAQIRPVTRINDSRVDVEYIDPVHPVYVIRINFTGNEKT 366
Db 301 VTRMEDDIKMLGRYGYAYPRVVTQPEINDDDKTVKLHINVDAGNRRFYVRHIREGNDTS 360
QY 367 QDEVLRREMRQEGGWASLYILIDQSKARLERLGKFEVNEVPANPGTDQDVVNSVE 426
Db 361 KDSVLRREMRQEGAWLGNDOVEAGKERLNLGYFEIVDVETQVRPGAADLVDVTKVKE 420
QY 427 QPSGSSTIAAGYSQSGVTFQFDVSONPMGTCKKHNASFSRSETRVYSLGNTDNYFT 486
Db 421 RNTGSLNFGIGYGTESGVSVQVQVQDNWLTGNTGVTGKNDYQTYAEFTLMDPYFTV 480
QY 487 NGVSQSLSGYYRKYTKYDNKNISNVYLDVSYGSSLSYGYPIDENQRIISGLNADNKLH 546
Db 481 DGVSLGRRIFTNDFKADNADLSGYTNSYAGDGLGFFPINENNSLRVGVYVHNDL-SDM 539
QY 547 FMGISNVKQLMADGGKIQVDNNGIPDFKHDTYTNAILGWNYSILDRPVFTQGMHSHVD 606
Db 540 LPQVAMWRYLESVGERPGYD--GREGFTTDDFTLN--LGWTYNLDRGFFFTSGYKSVN 595
QY 607 --LTVGFGDKTHQKVVYOGNIYRPF--IKKSVLRGYAKLGYGN-----NLPIYENFYAGG 657
Db 596 TKITVPGSDNEFYKVTFTDSAYQPLNEDRSVLLGRGLGYGDGIGSKEMPIYENFYAGG 655
QY 658 YGSRVGYDSSGLGPRSOAVLTARRGQQTTLGGEVVGGNALATFGSELIILPLPFKGD-WIDQ 716
Db 656 SSTVRGFRSNNIGPKA-AYANGGATVNTSDAVGNAVAVASIELITPTPFISEKYSNS 714
QY 717 VRPVIFIEGGQVDFDTGMDKQTDLTQFKDPQATASQNAKANRPLLTQDKQLRYSGAVG 776
Db 715 VRTSIFIDSGTWMDTNW-----ENTAK--TRAAGIPDYGKASNIRVSAGVA 758
QY 777 ATWYTPIGPLSISYAKPLNKKONDQDTVQFOIGSVF 813
Db 759 LQWMSPLGPLVFSYAKPVKDYEGDKSEQQFQFNIGTKW 795

RESULT 3
AB0530
outer membrane protein precursor yaer [imported] - Salmonella enterica subsp. enteric
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB0530
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Fair
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AB0530
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD08682.1; PID:gl6501505; GSPDB:GN00176
C:Genetics:
A:Gene: yaer
C:Superfamily: protective surface antigen D-15

Query Match 29.0%; Score 1218.5; DB 2; Length 803;
Best Local Similarity 33.9%; Pred. No. 1.2e-62;
Matches 285; Conservative 150; Mismatches 327; Indels 79; Gaps 15;
QY 14 MTMAVMMVMS-----THAQADFMANDITITGLQRTVIESLQSVLPFRGQVYSENOLA 67
Db 1 MAMKLLIASLLPSSATVYAGEGFVVDIHFEGQLQVAVGAALLSMFVRTGDTVNDDEIS 60

QY	68	DGVKALYATGNFSDVOYVTHOEGRIIYQVTERPLTAEINFGNRLIPKQEGELKNAGIA	127
Db	61	NTIRALFATGNFEDVRVLRDGNLTLLVQKERTPTASITFSGNKSVKDDMLKQKLEASGYR	120
QY	128	VGPQLKQATQMIETELTNOYISQYYNTEITVTKQTMDCNRVKLDMTAEKGPBARVDI	187
Db	121	VGESLDRITLSDIEKGLEDFYVSGKVSASVAVVITPLPNRVVDLKLIVQEGVSAKIQI	180
QY	188	NIIGNQHSFSDADLLDVLAKD-----NKINPLSKADR-YTQEKLVTSLENLAKYINAGF	241
Db	181	NIVGNHAFSTBELLSHFQLEDPVWNKV-----GDRKKQKQKLAGDLETLSRYVLDRGY	235
QY	242	VRFEIKDAKLINEDKNRIFEVLSHEGEQYRFQQTGLNLTNYTQAELEALLKPKAEGS	301
Db	236	ARFNIDSTQVSLTPDKGIYTVNITGEDQYKLSGVQVSGNLAGHSAETKLTKEPGL	295
QY	302	FSQAMLEQTTNNISTKFGDDGYYYAQIRPVTRINDESRITVDVEYYIDPVHPVYRRINFT	361
Db	296	YNGTKVTKMEDDIKLLGRGYAYPRVQSQPEINDADKTVKLRVNVDAAGNREVVRKIRE	355
QY	362	GNFTQDEVLRREKROLLEGALASNOKIQLSRALMTGFFKHYVTVDTRPVNSPDQVDV	421
Db	356	GNDFSKDSVLRRKRMQEGALSGDLVDQGERLNRLGFFETVDTTRQVPGSPQVDVY	415
QY	422	FVVEEQSGSSTTAAGYSQSGGYTFQFDVSONFMGTGKHVNASFSKSTREVISLGMTN	481
Db	416	YKVKERTGTFNFGIGYGTESGVSFQAGVQDDNWLGTGYSVGINGTKNDYQIYSELSTN	475
QY	482	PYFTVNGVSQSLSGYYRKTYYDNKNISNVYDLSYGSLSYGYPIDENQRISFGLNADNTK	541
Db	476	PYFTVDSVSGGRIFYNDFEADDADLSDVTKNSYGTVDTLGPPINEYNLRAGLGVVHNK	535
QY	542	LHGGREMGISNVK-OLMADGGKIQVDNNGIPDPFKHDYTYTNAILGNWYSSLDPRVPFTQG	600
Db	536	L-----SNMQQITAMDRYLESMGDPDASDFAADDFTFN--YGWTYKNKLDGFYPTDG	585
QY	601	--MSHSYDLTVGFGDKTHQVYVQGGNIYRPFIR--KSVLRGYAKLGYGNL-----PFVE	651
Db	586	SRVNLTKVITPGSDNEYKYVSLDTATYVPIDNDHKWVVLGRTRWGYDGLGCKEMPFFE	645
QY	652	NFYAGGYSVRGYDQSSLSGPRSQAYLTARGQOTT-----LGEVVGGN	694
Db	646	NFYAGGSSTVRGFSQNTIGPKA---VYKNGAHTSWDDDDYEDCTQESGCKSDSDAVG	701
QY	695	ALATFGSELILPLPFKGD-WIDQVRVPIFTEGGQVFDTT-GMDKQITDITQFKDQATAE	752
Db	702	AMAVASLEFITPPTFISEKANSVRFSFFWDGXTVWDTNWDSSAPSDVPDYSDP-----	756
QY	753	QNAKAANRPLLTOOKLRSYAGVATWYTPIGPLSISYAKPLNKKONDQTDTPVQFQISV	812
Db	757	-----GNIRMSAGIALQWMSPLGLPVFSYAQPFKKYDGDGKAEQFQFNICKT	802
QY	813	F 813	
Db	803	W 803	
RESULT	4		
A64742			
hypothetical protein b0177 - Escherichia coli			
C:Species: Escherichia coli			
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999			
C:Accession: A64742			
C:Author: Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.;			
C:Reviewer: D.J.; Mau, B.; Shao, Y.			
C:Complete genome sequence of Escherichia coli K-12.			
C:Accession: A64720; MUID:97426617			
C:Date: 1453-1462, 1997			

sequence not shown; translation not shown

A: Cross-references: GB: AE000127; GB: U00096; NID: g1786370; PIDN: AAC73288.1; PID: g1786370
 A: Experimental source: strain K-12, substrain MG1655
 A: Superfamily: protective surface antigen D-15

Query Match	28.8%;	Score 1210;	DB 2;	Length 810;
Best Local Similarity	34.0%;	Pred. No. 3.9e-62;		
Matches 287;	Conservative 149;	Mismatches 330;	Indels 78;	Gaps 15;

[illegible]

RESULT 5

C50631 hypothetical protein ECs0179 [imported] - *Escherichia coli* (strain 0157 H7, substrain

Db 706 VGNAMAVASLEFPTTPTFFISDKYANSVRTSFFWDMGTVMWDTNWDSSQYSGYPDSDP-- 763

QY 750 TAEQNAKAANRPLLTQDKOLRYSAGVATWYTPIGPLSISYAKRPLNKKQNDQTDVQFI 809

Db 764 -----SNIRMSAGIALQWMSPLGLVFSYAQPFKKYDKGDKABQFQNI 806

QY 810 GSVF 813

Db 807 GKTW 810

RESULT 6

C85502

hypothetical protein yaeR [imported] - Escherichia coli (strain O157:H7, substrain ED

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-feb-2001 #text_change 14-Sep-2001

C:Accession: C85502

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: C85502

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-810 <STO>

A:Cross-references: GB:AE005174; NID:g12512902; PIDN:AAG54479.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yaeR

C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;

Best Local Similarity 34.0%; Pred. No. 3.9e-62;

Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

QY 14 MTAVMMVMS-----THAQADFMANDITITGLQRTVIESLSQVLPFRGQVVSQNLA 67

Db 1 MAMKLLIASLFFSSATVYGAEGFVVKDIHFEGQLQAVGAALLSMFVRTGDTVNDDEIS 60

QY 68 DGKALYATGNFSDVQVYHOGRIIYOVTERPLIAEFNFGNRLIPKEGLEGLKNAGLA 127

Db 61 NTRALFATGNFEDVRLRGDITLLVQKERTIASITFSGNKSVKDDMLKNLEASRG 120

QY 128 VGQPLKQATQVMIETELTNQYISQYNTETITVKTMLDGNRVRKLDMTFAEGKPARVVDI 187

Db 121 VGESLDRITTIADIEKGLDFYYSVGKYSASVAVVTPLPNRNVDLKLVFQEGVSAEIQI 180

QY 188 NIIGNQHFSDADLLDVLAIKD-----NKINPLSKADR-YTOEKLVTSLNLRKYLNAGF 241

Db 181 NIVGNHAFITDELISHFQLRDEVPWNVV-----GDRKYQKQKLAGDLETLSRYLDRGY 235

QY 242 VRFEIKAKLINEDKKNRIFVEISLHGEQYRFQGTQFLGNLTQVTOAELEALLKFAEEG 301

Db 236 ARFNIDSTQVSLTPDKKGIYVTVNITGEDQYKLSGEVSGNLGASHAEIQLTKEPGE 295

QY 302 FSQAMLEQTNNISITKFGDDGYIAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFT 361

Db 296 YNGPKVTKMEDDIKLLGRYGYAYPRVQSMPEINDADTKLVKNVNDAGNRFYVRKIRFE 355

QY 362 GNFTQDEVLRREMRQLEGALASQKIQLSARLMRTGFFKHVTVDFRPVNSPDQVDV 421

Db 356 GNDTSKDAVLRREMRQLEGALASQKIQLSARLMRTGFFKHVTVDFRPVNSPDQVDV 415

QY 422 FVBEQPSGSGSTIAAGYSQSGGVTFQFVSONNPMGTGKHVNASFSSRSETREVYSLGMTN 481

Db 416 YKVERNTGSENFNGIGYTESGVQFAGVQDQWNLGTGYAVNGTNGDYTYAELSVTN 475

QY 482 PYFTVNGVQSLSGYRKTKYDNKNISNYLDSYGGSLSGYGPIDENQRTISFGLNADNTK 541

Db 476 PYFTVDGVSGLGRIFYNDFQADADLSYTNKSYGTDVTLGFPINEXNSLRAGLGYVHNS 535

QY 542 LHGRFPWGISNVKQLMA-----DGGKIQVDNNGIPDFKHDTYTNALLGNYSLSLRPV 595

Db 536 L-----SNMQPOVAMRWLYLSMGEHPSTSDQDNSFKTDFTFN--YGMTYKNLDRGY 585

QY 596 FPTQG--MSHSVDLTGFGDKTHQKVYVQGNIRPF--IKKSLRGYAKLGYGNL----- 647

Db 586 FPDGSRVNLTKGTIFGSDNEYKVTLDATYVPIDDDKHWVLGRTRWGYGDLGKGE 645

QY 648 -PFYENYAGGYGVRGVDQSSLGRS-----QA-----YLTARRGQQTTL---GEV 690

Db 646 MPFYENYAGGSSTVRGFSQNTIGFKAVYFPHQASNDPDYDYECATQDGAKDLCKSDDA 705

QY 691 VGNALATFGSELILPLPFKGD-WIDQVRPVIFTEGGQVDFTTGMDKQTDILTQFKDQPA 749

Db 706 VGNAMAVASLEFPTTPTFFISDKYANSVRTSFFWDMGTVMWDTNWDSSQYSGYPDSDP-- 763

QY 750 TAEQNAKAANRPLLTQDKOLRYSAGVATWYTPIGPLSISYAKRPLNKKQNDQTDVQFI 809

Db 764 -----SNIRMSAGIALQWMSPLGLVFSYAQPFKKYDKGDKABQFQNI 806

QY 810 GSVF 813

Db 807 GKTW 810

RESULT 6

C85502

hypothetical protein yaeR [imported] - Escherichia coli (strain O157:H7, substrain ED

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-feb-2001 #text_change 14-Sep-2001

C:Accession: C85502

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: C85502

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-810 <STO>

A:Cross-references: GB:AE005174; NID:g12512902; PIDN:AAG54479.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: yaeR

C:Superfamily: protective surface antigen D-15

Query Match 28.8%; Score 1210; DB 2; Length 810;

Best Local Similarity 34.0%; Pred. No. 3.9e-62;

Matches 287; Conservative 149; Mismatches 330; Indels 78; Gaps 15;

QY 14 MTAVMMVMS-----THAQADFMANDITITGLQRTVIESLSQVLPFRGQVVSQNLA 67

Db 1 MAMKLLIASLFFSSATVYGAEGFVVKDIHFEGQLQAVGAALLSMFVRTGDTVNDDEIS 60

QY 68 DGKALYATGNFSDVQVYHOGRIIYOVTERPLIAEFNFGNRLIPKEGLEGLKNAGLA 127

Db 61 NTRALFATGNFEDVRLRGDITLLVQKERTIASITFSGNKSVKDDMLKNLEASRG 120

QY 128 VGQPLKQATQVMIETELTNQYISQYNTETITVKTMLDGNRVRKLDMTFAEGKPARVVDI 187

Db 121 VGESLDRITTIADIEKGLDFYYSVGKYSASVAVVTPLPNRNVDLKLVFQEGVSAEIQI 180

QY 188 NIIGNQHFSDADLLDVLAIKD-----NKINPLSKADR-YTOEKLVTSLNLRKYLNAGF 241

Db 181 NIVGNHAFITDELISHFQLRDEVPWNVV-----GDRKYQKQKLAGDLETLSRYLDRGY 235

QY 242 VRFEIKAKLINEDKKNRIFVEISLHGEQYRFQGTQFLGNLTQVTOAELEALLKFAEEG 301

Db 236 ARFNIDSTQVSLTPDKKGIYVTVNITGEDQYKLSGEVSGNLGASHAEIQLTKEPGE 295

QY 302 FSQAMLEQTNNISITKFGDDGYIAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFT 361

Db 296 YNGPKVTKMEDDIKLLGRYGYAYPRVQSMPEINDADTKLVKNVNDAGNRFYVRKIRFE 355

QY 362 GNFTQDEVLRREMRQLEGALASQKIQLSARLMRTGFFKHVTVDFRPVNSPDQVDV 421

Db 356 GNDTSKDAVLRREMRQLEGALASQKIQLSARLMRTGFFKHVTVDFRPVNSPDQVDV 415

QY 422 FVBEQPSGSGSTIAAGYSQSGGVTFQFVSONNPMGTGKHVNASFSSRSETREVYSLGMTN 481

Db 416 YKVERNTGSENFNGIGYTESGVQFAGVQDQWNLGTGYAVNGTNGDYTYAELSVTN 475

QY 482 PYFTVNGVQSLSGYRKTKYDNKNISNYLDSYGGSLSGYGPIDENQRTISFGLNADNTK 541

Db 476 PYFTVDGVSGLGRIFYNDFQADADLSYTNKSYGTDVTLGFPINEXNSLRAGLGYVHNS 535

Qy 542 LHGGRFMSISNVKQLMA-----DGGKIQVDNNGIPDPKHDYTTYNAILGNWYSSLDPRV 595
Db 536 L-----SNMQPQVAMWRYLSMGEHPSTSDQNSFKTDDFTFN--YGTYNKLDGRY 585
Qy 596 FPTQO--MSHSVDLTNGFGDKTHOKVYVOGNIYRPF--IKKSVLRGYAKLGYGNL----- 647
Db 596 FPDGSRVNLTKVTIPGSDNEYKYVTLDTATYVPIDDHKKWVLGRTRWYGGDLGKGE 645
Qy 648 -PFYENFYAGGYGVRGVDQSSLGPRS-----QA-----YLTARRGQOQTTL---GEV 690
Db 646 MPFENFYAGGSSTVRGFSQNTIGPKAVYFPHQASNYDPDYDECATQDGAKDLCSDDA 705
Qy 691 VGGNALATFGSELILPLPFFKGD-WIDQVRPVIFIEGGQVFTTGMKDKTIDLTQFKDPQA 749
Db 706 VGGNAMAVALFEFTPTFFISDKYANSVSTFFWDMGTWDTNWDSSQYSGYPDYSDP-- 763
Qy 750 TAEQNAKAANRPLLTQDKQLRYSGAGVATWTPITGPIGLSISYAKPLNKKONDOTDTVQFOI 809
Db 764 -----SNIRMSAGIALQWMSPLGLVFSYAQPFKKYDGDKAQFQFQNI 806
Qy 810 GSVF 813
Db 807 GKTW 810

RESULT 7
B82099
surface antigen VC2252 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: B82099
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vanathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, H.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82099
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <HEI>
A:Cross-references: GB:AE004297; GB:AE003852; NID:9656810; PIDN:AAF95396.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC2252
A:Map position: 1
C:Superfamily: protective surface antigen D-15

Query Match 28.6%; Score 1203; DB 2; Length 803;
Best Local Similarity 33.4%; Pred. No. 9.8e-62;
Matches 278; Conservative 155; Mismatches 328; Indels 72; Gaps 15;

Qy 16 MAVVMVWSTHAQAD-FWANDITITGLQRTVIESLQSVLPRLGQVSVENQADGVKALY 74
Db 8 LATLATSVSANGAEKFVQDIQDGLGVALGAALKMLPVRVGSDVSDQVANIIRALY 67
Qy 75 ATGNFSDVQVYHQBGRYIYQTERPLIAEFNFEGRNLPIKREGLQGLKNAGLAVGQPLKQ 134
Db 68 SSGNFEDVKVLRDGNLTLMVQVKERTIASVSFSGNKAKEBQLQNLQEAASSIRVEALDR 127
Qy 135 ATVOMIETELTNQYISQYNYETVVTQMLDGNRVKLDWTFAGKPARVVDVIMINQNH 194
Db 128 TTLSNIEKLEGLDFYISGKYNATKAVVVTPLPRNRADLKFVFTGVSAAKIQINFINQV 187
Qy 195 FSDADLLDVLAIKDNKI--NPLSKADRYTOEKLVTSLNENRAKYNLACGVRFEIKDAKLN 252
Db 188 FSDIELSRFLNVDVAVWNNFLAD-DKYKKQVLADGIEALRTYYLDRGYLKFQVDSTQVA 246
Qy 253 INEDKNRIFVSLHEGEQYRFQTOFLGNLTYYQAELEALLKFKAEFGFSQAMLEQTTN 312
Db 247 ISPDKGVYITLNLNEGEPTVSKVQFGLMGKAEFTSLIPEIGETNGSAVTRLEE 306

Qy 313 NISTKFGDDGYYIAQIRPVTRINDESRTVDVEYYIDPVHVVVRRINFTNFKTKQDEVLR 372
Db 307 SVKKVLGSEYAYPOVRTIPEFDEKQCOQSLVHVVEACKRVVVRDIRFVGNNSTRDEVLR 366
Qy 373 REMQLEGALASNQIKLSRARLMTGTFKHKVTVTRVPVNSPDQVDVNFVVEEQSGSS 432
Db 367 REMQMEGSLWNLGSDIETGKTRNLRLGFFETVETVQVRVPGSEDQVDLVYSYKEANSNV 426
Qy 433 TIAAGYSGSGVTFQFDVYSONNFMTGTRKHVNASPFSRSETREVYSLGTMNPNPTVTVNGVSOS 492
Db 427 NFGVGYGTESGVSTQVGLQDNFLGSGNRVGVNAMINDYQKNLTLEYRDPYWNLDGVSIG 486
Qy 493 LSGYYRKYTKYNKKNISNVLDYSYSGSLSYGYPIDNQRIISFGLNADNTKLHGGRPMGTSN 552
Db 487 GKVEYNQPEASEAGIVDTNESYGTSLTWGFPDFELNRFEGIGYTHNKI--GNLTPIYLQ 544
Qy 553 VKQLMA-----DGGKIQVDNNGIPDPKHDYTTYNAILGNWYSSLDPRVPTQGMHS- 604
Db 545 VENFLAAQASNIDSGNLLTDD-----FDINLSWTRNNLNNSYFPTAG-NHQ 591
Qy 605 --VDLTGFGDKTHQKVYVYQNIYRPFIKKS-----VLRGYAKLGYGN-----LP 648
Db 592 AFYKMTVPGSDAQYFKLOYDVROYEPLTKKHEFTLLLRG--RLGYNGYGGQTDGKDNLEP 649
Qy 649 FYENFYAGGYGSRVGYDQSSSLGPRS--QAYLTARRGQOQTTLGEVVGGNALATFGSELILP 706
Db 650 FYENFYAGGFTSLRFGSGSAGPKAVRYDYSNGNSDTSATDDSVGGNAIALASVELIVP 709
Qy 707 LPFKGDWI-DOVRPVIFIEGGQVFTTGMKDKOTID-----LTQFKDPQATAEONAKAANR 760
Db 710 TPFASEARNQIRTSIFYDMAVMDTEFDYRGKADYGNQYVYDSDP----- 756
Qy 761 PLLTQDKQLRYSGAGVATWTPITGPIGLSISYAKPLNKKONDQTDVTFQIGSVF 813
Db 757 -----TNYRSSYVALQWSPGMLPVFLSLAKPIKKYEGDDEEFTTIGRTF 803

RESULT 8
E82731
outer membrane antigen XF1046 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: E82731
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: E82731
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-784 <STM>
A:Cross-references: GB:AE003941; GB:AE003849; NID:9105978; PIDN:AAF83856.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferrelira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1046
C:Superfamily: protective surface antigen D-15

Query Match 27.9%; Score 1172.5; DB 2; Length 784;
Best Local Similarity 33.8%; Pred. No. 5.5e-60;
Matches 278; Conservative 151; Mismatches 321; Indels 73; Gaps 13;

QY 14 MTAVMVMVSTHAQAAD-FMANDDITITGLQRTVIESLQSVLPFLQGVVSENQADGKA 72
DB 12 LTAANFSLPVLQAASFVANDIRVDGLQRIASGVFTYLPVNRGDTVDKADVAIRA 71
QY 73 LYATGFSDDQVYHQEGRIIVQTERPLIAEINFEQNRLLPKEGLOEGLKNAGLAVGQPL 132
DB 72 LYRTGFENVIRDRQGNILVVKVERPAINKLITGKNKDKSEELLKGLSEIGLSEGGTF 131
QY 133 KQATVQMIEFELNXYISQGYNTEITVQTMLDGNRVKLDMTFAEKPAPRVVDINIIGN 192
DB 132 DRLSLDRVTQELKROYNRRKGNVQMTTTPLDLRNRVDVTIAKEGKAARHINLIGT 191
QY 193 QHESDADLIDVLAIKD-NKINPLSKADRYTOEKLVTSLNLRKAYLNAGVREIKDAKL 251
DB 192 EKENNDKVMGAWSEKHNWASWYRDDQYSKEKLSGDKLNSWYLDRGVYDFNIDSTQV 251
QY 252 NINEDKNRIEVEISLHEGEYRFQOTQFLGNLTYYTOAELEALLKFAEEGFSQAMLEQTT 311
DB 252 SISPEKHMFITAGVTEGQYKISSIKVTGNTVLPQEKIEKLVIPKTDGIFSRVLEYSS 311
QY 312 NNISTFGDDGYYYAQIRPVTRINDESRVDEYIIDPVHPVYVRINFTGNPKTQDEVL 371
DB 312 AAILNTLSNIGYAFSKYNPITPTANRADRTVAVNLHVIPGRVTVRQILFKGNTRTSDEVL 371
QY 372 RREMROLEGALASNOKLTOLSRARLMTGFPKHVTVDTRPVNPSDQVDVNFVVEEQPSGS 431
DB 372 RREMROFENSWYSOAALDRSKIRLQRLGYFEADVDESTVPVGSNDQVDIIVYTKETSGS 431
QY 432 STIAAGYSQGGVTFQFDVSNQNMFGTGHVNASFSRSETREYVSLGNTNPIYFTVNGVSQ 491
DB 432 FOYGLGYSKYTVTTVOLSONNFGLSGNRVSDASRSRYQDRYSPSYNPNPFTDNGVSL 491
QY 492 SLGYSRKTKYDNKNTSNVYLDYSGSLSYGYPIDENQRISFGLNADNTKLHGGRFMGIS 551
DB 492 GYNLAYOKLDYDFNAAQYNSKRMSGQTFIGIPITENDTVSWVIGADSNQI--TTPPG-S 548
QY 552 NVKOLMADGGKIOVDNNGIPDFKHDTTYNAIILGNYSLSLDRPVFPQGMHSHVLDLVGTF 611
DB 549 TPKRAID-----YIDAVG-----QRTFAMWTELGWARDTRNDYFMPNLGMQIRIGAEVTL 599
QY 612 GDKT--HQKVYVQGNIRPPIKKSRLGYAKLGYGN-----LPHY 650
DB 600 PGSTIKYKINYQISKYWPILPALVNLRLVGVGDDYGKSHRILPDGTVATASGLPFF 659
QY 651 ENFYAGYGSVRGYDOSSIGPRSOAYLTARRGOQTLTGEVVGNGALATFGSELILPLPFK 710
DB 660 ENFYAGTNSVRGFRONTLGRSE--VTALYNQ---GQPLGGSFKTVGSGTEMYFPKLEF 713
QY 711 GDWIDQVRPFIETEGGOVFTTGMDKQTLDTQFKDPOATAEQNAKAAANRPLLTQDKOLR 770
DB 714 S---PSARISAFIDFGVFN-----GVNFKA-----NELR 741
QY 771 YSAGVGATWTPICPLSISYAKPLNKQNDQTDVTFQIGSVF 813
DB 742 ASSGVALLWRAPICPISISYAFPIKKNENDEIERLQFTFGQGF 784

RESULT 9
G81228
outer membrane protein Omp85 NMB0182 [Imported] - Neisseria meningitidis (strain MC58 se
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81228
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A:Reference number: A81000; MUID:20175755
A:Accession: G81228
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-797 <DET>
A:Cross-references: GB:AE002375; GB:AE002098; NID:g7225394; PIDN:AAF40639.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB0182
C:Superfamily: protective surface antigen D-15

Query Match 26.5%; Score 1115; DB 2; Length 797;
Best Local Similarity 32.5%; Pred. No. 1.2e-56;
Matches 266; Conservative 154; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTMAVMVMVSTHAQAADFMANDITITGLQRTVIESLQSVLPFLQGVVSENQADG 69
DB 2 KLKOIASALMMLGISPLALADFTIQDIRVEGLQRTSTVFVNLVVKVGDYNDTHGSAI 61
QY 70 VKALYATGNFSDDQVYHQEGRIIVQTERPLIAEINFEQNRLLPKEGLOEGLKNAGLAVG 129
DB 62 IKSXYATGFPDDVRVETADGQLLLTVIERPTIGSLNITGAKMLQNDAIKKNLESFGLAQS 121
QY 130 QPLKQATVQMIEFELNXYISQGYNTEITVQTMLDGNRVKLDMTFAEKPAPRVVDINI 189
DB 122 QYENQATLNOAVAGLKEEYLGRLKNIQITPKVTKLARNRVDITITDEGSAKITDIEF 181
QY 190 IGQHESDADLIDVLAIKDNKI-NPLSKADRYTOEKLVTSLNLRKAYLNAGVREIKD 248
DB 182 EGNQVSDRKLMDQMSLTGEGITWLTSTRNOFNEQPAQDMKEKVTDFYQNGVDFDFRILD 241
QY 249 AKLINEDKNRIEVEISLHEGEYRFQOTQFLGNLT-YTOAELEALLKFAEEGFSQAML 307
DB 242 TDLOTNEDTKQIKITVHEGGRFNGKVSIEGDNEVPKAELEKLLTMPKQWYERQOM 301
QY 308 EOTNNISTKFGDDGYYYAQIRPVTRINDESRVDEYIIDPVHPVYVRINFTGNPKTQ 367
DB 302 TAVLGEIQNRMGSGAYSEISVQPLPNAETKTVDVFLHIEPGRKIVYNEIHTGNKTR 361
QY 368 DEVLREMRQLEGALASNOKLTOLSRARLMTGFPKHVTVDTRPVNPSDQVDVNFVVEEQ 427
DB 362 DEVVRELQMSAPYDTSKLQSKERVELLYGFDVNFQFADVPLAGTDPKVDLNMSTER 421
QY 428 PSGSSTIAAGYSQGGVTFQFDVSNQNMFGTGHVNASFSRSETREYVSLGNTNPIYFTVN 487
DB 422 STGSLDLSAGVQDTGLVMSAGVSQDNLCFCTGKSAALRASRKTLLNGSLSTDPYFTAD 481
QY 488 GVSQSLSGYYR--KTKYDNKNTSNVYLDYSGSLSYGYPIDENQRISFGLNADNTKLHGG 545
DB 482 GVSGLGYDVYKAFDPRKASTSIKQYKTTTAGAGIRMSVPTVTEYDRVNFGLVAEHLATVN-- 539
QY 546 RFMGISNVKOLMADGGKIOVDNNGIPDFKHDTTYNAIILGNYSLSLDRPVFPQGMHSHV 605
DB 540 ---TYNKAPKHYADFIKKYKKTG--TDGSPKGLYKGTGVGWRNKTDALWPRGYLTGV 595
QY 606 DLTGVF--GDK--THQKVYVQGNIRPPIKKSRLGVLR-----GYAKLGY--NNLPYENFYAG 656
DB 596 NAEIALPGSKLQIYSATHNQNTWFFPLSKTFTLMLGGEVGIAG--GYRGTKEIPFENFYAG 654
QY 657 GYGSVRGYDOSSIGPRSOAYLTARRGOQTLTGEV--GGNALATFGSELILPLPFKDWI 714
DB 655 GLGSVRGYESTGLPK-----VYDEYGEKISYSGNKKANVSALLPMPFGAKD-A 703
QY 715 DOVRPVFIETEGGOVFTTGMDKQTLDTQFKDPOATAEQNAKAAANRPLLTQDKOLRYAS 774
DB 704 RTVRLSLFADAGSVWDGKTDYDDSSSAT-----GGRVONIYGAGNTHKSTFTNELRSAG 758
QY 775 VGATWTPICPLSISYAKPLNKQNDQTDVTFQIGSVF 813
DB 759 GAVTWLSPLGPMKFSYAYPLKKKPEDEIORFQOLGTF 797

RESULT 10
D82000

outer membrane protein OMP85 NMA0085 [imported] - Neisseria meningitidis (strain 22491 s
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: D82000
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
S.; Holroyd, S.; Jagerski, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: A81775; MUID:2022556
A:Accession: D82000
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-797 <PAR>
A:Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83401.1; PID:g737885
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: omp85; NMA0085
C:Superfamily: protective surface antigen D-15

Query Match 26.4%; Score 1111; DB 2; Length 797;
Best Local Similarity 32.4%; Pred. No. 2e-56;
Matches 265; Conservative 155; Mismatches 361; Indels 38; Gaps 14;

QY 10 QVSAMTMAVMVASTHAQAADFMANDITITGLQRTTIESLQSVLPFRLLGQVWSENQIADG 69
DB 2 KIKQIASALMVLGISPLALADFTIQDIRVEGLQRTSTVFNVLPVKVGDYNDTHGSAI 61
QY 70 VKALIATGNFSDVOVYHQEGRIYQVTERPLIAINEFGRNLIPKQGLGLKNAGI 129
DB 62 IKSLYATGFFDDVRVETADQGLLTIVTERPTIGSLNITGAKMLQNDAIKNLESFGLAQS 121
QY 130 QPLKQATQVQMEITELTNQYISQGYNYETITVKTMDLGNRVKLDMTFAEGKPARVVDINI 189
DB 122 QYFNQATLQNAVAGLKEEYLGGRKLNQITQPKVTKLARNRVDIDITIDEKSAKITDIEF 181
QY 190 IGNOHFSDADLDIDVIAITKDINKI-NPLSKADRYTQEKILVTSLLENRAKYLNAAGFVRFEIKD 248
DB 182 EGNQVYSDRKLMRMSLTGEGIIWTLTRSNQFNQEKQADMEKVTDFYQNNGYDFDRILD 241
QY 249 AKLNINEDKNRIFVEISLHGEQYRFQOTQGLNLT-YTQAELEALLKFAEGFQSAML 307
DB 242 TDIQTNEDTKTKQTKITVHBEGRPWGKSTEGDTEVPKAELEKLLTMKPGKWERQOM 301
QY 308 EQTNNISTFGDDGYAQAIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNKQTQ 367
DB 302 TAVLGETONRMGSAGYAYSEISVQPLPNAETTKTVDFLHIEPGRKIYVNEIHTNNKTR 361
QY 368 DEVLRRMROLEGALASNQIKQLSRALMRGTGPFKHVTVDTRPVPSNPDOVDVNFVVEEQ 427
DB 362 DEVVRELROMESAPYDTSKLQSRKERVLLGYFDNVQFQDAVPLAGTPDKVDLNMSTLR 421
QY 428 PSGSSTIAAGYSQGGVTFQFDYSQNNFMGTGKHVNASFGRSRETREYISLGMTNPFYTN 487
DB 422 STGSLDLSAGWQVDTGLVMSAGVSODNLFTGTGKSAALRASRSKTTLNGSLSFDPYETAD 481
QY 488 GVSOSLSGYR--KTKYDKNKINSYVLDSGGSLSYGYPDENORISFGLNADNTKLHGG 545
DB 482 GVSIGYDVYKAFDPRKASTSIKQYKTTAGAGIRMSVPVTEYDRVNFVAEHLTV-- 539
QY 546 RFMGISNVKQLMADGGKIQVDNNGIIPDKHDYTYNAILGWNYSLLDRPVFPPTGCMHSV 605
DB 540 ---TYNKAPHYADFIKKYKGTG--TGGSFKGWLYKGTVMGRNKTDSALMPTKGYITGV 595
QY 606 DLTVGF-GDK--THOKVYVYQGIYRPFPIKKSVLK-----GYAKLGCG--NNLPFYENFYAG 656
DB 596 NAEALPGSKLOYYSATHNTQWTFPFLSKTFLMLGGEVGING-CYGRTKELIPFFENFYGG 654
QY 657 GYGSVRGYDSSLGPRSOAYLTARRGOQTTLGVEV--GGNALATFGSELILPLPFPKGDWI 714
DB 655 GLGSVRGYESTGICPK-----VYDEYGEKISYGGNKKANVSAELLPPMPGAKD-A 703

Qy	715	DQVRPVIFEGGVDDTTGMDKQTIDLTQFKDPOATBAONAKANRPLLTQDKQLRSYAG	77
Db	704	RTVRLSLFADAGSVWDGKTYYDNSSSAT-----GGRVNITYGAGNTHKSTFTNELRYSAG	758
Qy	775	VGARWTPPIGLPSISYAKPLNKKNODGTDTVQFOIGSVF	813
Db	759	GAVTWLSPGLPMKFESYAYPLKKPKPEDEIQRFOQLGTTTF	797
 RESULT 11 F64102 protective surface antigen D-15 - Haemophilus influenzae (strain Rd KW20) C:Species: Haemophilus influenzae C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1999 C:Accession: F64102 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F. J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J. . D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geogha Science 269, 496-512, 1995 A:Authors: Gnehm, C.U.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, A::Title: Whole-genome random sequencing and assembly of Haemophilus influen A:Reference number: A64000; MUID:95350630 A:Accession: F64102 A>Status: nucleic acid sequence not shown; translation not shown A:Molecule type: DNA A:Residues: 1-808 <TIGR> A:Cross-references: GB:L42023; TIGR:HI0917 C:Superfamily: protective surface antigen D-15 C:Keywords: surface antigen			
Qy	23	STHAAQAFDMANDITITGLQRTVTIESLSQVLPFLPGQWSENOLADGKALYATGNFSDV	82
Db	27	TTFVFAAPPFAKDIRVDGVQGDLEQQIRASLPVRAGORVTDNDVANIVRSLSFGRFDDV	86
Qy	83	QVYHOEGRI-IYQVTERPLIAINFEGNRILPKEGLQELKKNAGLVGPQLKQATVMIE	1414
Db	87	KA-HQEGDLVVSVVAKSIISDVIKGNSIIPTALKQNLDANGFKYGDVLIREKLNEFA	145
Qy	142	TELTNQYISQGYNYTEITVKQTMLDGHRVKLDMTFAEKGPARVVDINIIGNQHFSADLI	201
Db	146	KSVKEHYASVGRYNAIVEPIVNTLPNNRAEIIQINEDDKAKLASLTFFKGENSVSSTLQ	205
Qy	202	DVLAIKDNKINPLSKADRYTOEKLVTSLENLRKYLNAGVFPRFIKDAKLININEDKNRIF	261
Db	206	EOMELQDPDSWWKL-WGNKFCAQAEKDQLSIRDYLLNGYAKAQITKTDVQLNDEKTKVN	266
Qy	262	VEISLHEGEQYRGOTQFQLGNLTVTAOLEALLK-FKAEGBFSQAMLEQOTTNNISTWKFD	320
Db	265	VTIDVNEGQLDLARSARIIGMLGMSALEPLLSALHLNDTFRRSSDIADVENAIKAKLGE	324
Qy	321	DGYYYAQIPRYTRINDESRTVDVEYIDPVHPVYVRINFTGNFKTEODEVLRREMRLEG	380
Db	325	RGYGSATVNSVPDFDDANKTLAITLVDDGRRLLTVQRLREFGNTVSADSTLRQMRQOEG	384
Qy	381	ALASNQIKQLSRARLMRTGFPPKHVTVDTR--PVPNSPDQVDVNFVVEQPSSGSTIAAGY	433
Db	385	TWYNSQVELCKIRLDRTGFE--TVENRIDPINGSNDVDVYKVKERWTSGINFYGIGY	442
Qy	439	SOSGGVTFQFDVQNNFMFGTKGHWNASFERSRETREVYSLGMNPYFTVTVNGVOSLSGYR	498
Db	443	GTESISYQASVKQDNFLGTGAAYSIAGTKNDYCTSNVLGYTEPYFTKDG/-SLGNGNVF	500
Qy	499	KTKYDNK---NISNVYLDSYGGSLSYGPIDENQRISEGL-----NADNTKLHGCRFWGI	550
Db	501	FENYDNSKSDTSNNKRTTYGSNVTGLFPVNNENSYYVGLGHTYNKYISNFALEYNRNLXI	560
Qy	551	SNVKOLMADGGKIOVDNNGIPDFKHDTYTYNAILGWNYSSLDRPVPFTQGGHSHSD--LT	608

```
Db 561 QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYFPFKGVKASLGRGRT 605
Qy 609 VFGDKTHQKVVYQGNIRPF-----IKKSVLRGKYAKLGYGN-NLPFFYENFYAGGYGS 660
Db 606 IPGSDNKYYKLSADVOGFYPLDRDLHWVWSAKASAGYAN-GFGNKRLPFYQTYTAGGIGS 664
Qy 661 VRGYDQSSLSRPSQAYLTARGQOT---TIGEVVGGNALATFGSELILPLPFKGDWI-DQ 716
Db 665 LRGFAYGSIGPNA-IYAEHNGNGTFFKKISSDVIGGNAITATASAEELIVPTPFVSDKSQNT 723
Qy 717 VRPVIFIEGGQVFTT-GMDKQITDLTQFKDPOQATAEONAKAANRPLLTDQKQLRYSAGV 775
Db 724 VRTSLFVDAASVWNTKWKSDKSLD-----NNVLKSLPDYGGKSSRIRASTGV 770
Qy 776 GATWYTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813
Db 771 GFQWQSPIGLVFSYAKPIKKYENDVDEQFQFISIGGSF 808
RESULT 12
JC4078
protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
A:Variety: type b
C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C:Accession: JC4078
R:Flack, F.S.; Loomore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus influenzae
A:Reference number: JC4078; MUID:95255676
A:Accession: JC4078
A:Molecule type: DNA
A:Residues: 1-797 <FLA>
A:Cross-references: GB:U13961; NID:g537447; PIDN:AAA85645.1; PID:g537448
A:Experimental source: type b
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 25.08; Score 1052.5; DB 2; Length 797;
Best Local Similarity 31.68; Pred. No. 4.9e-53;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

Qy 23 STHAQAADFMANDITITGLQRTVTTIESLQSVLPFRLLGQVWSENQADGVKALYATGNFSDV 82
Db 14 TTTVFAAPFYAKDIRVDGVQGDLEQQIRASLPVRAGQRTDNDVANIVRSILFVSGREDDV 73

Qy 83 QVYHOGRI-IYQVTERPLIAEINFEGRNRLIPKEGLQGLKNAGLAVGQPLKQATVQMIE 141
Db 74 KA-HQEGDLVLSVVAKSIISDVKIKGNSVIPTALKONLDANGFKVGDVILIREKLENEFA 132

Qy 142 TELTNQYISQGYNTTEITVKTMLDGNRVKLDMTFAECKPARVVDINLIGNQHFSDADLI 201
Db 133 KSVKEHVASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASLTFKGNESVSSSTLQ 192

Qy 202 DVLAIKONKINPLSKADRYTOEKLVTSLENRAKYLNAGFVRFEIKDAKLNINEDKNRIF 261
Db 193 EQMELQPSWKKL-WGNKFEGAEQEKDLOSIRDYLLNNGYAKAOITKTDVQLNDEKTKVN 251

Qy 262 VEISLHEGEQYRFQGTQFLGNLTQTAELALK-FKAEEGFSQAMLEQTTNNISTKFGD 320
Db 252 VTIDVNEGLQDLRSARTIIGNLGMSAELEPLLSALHLNDTFRRSADIADVENAIKAKLGE 311

Qy 321 DGYYYAQIRPVTRINDESRSTDVEYIDPVHPVVRNFTNFTGNFKTODEVLRRMROLEG 380
Db 312 RYGSATVNSVPDDEANKTKLTLVLDAGRRLLTVROLRFEFGNTVSADSTURQEMROEG 371

Qy 381 ALASNQKIQLSRARLMRTKFKKHVTVDTR--PVPNSDQVDVNFVWEEQPSGSSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRTGFFE--TVENRIDPINGSNDEVDVYKVRNRTSGNSINFGIGY 429
```

```
Qy 439 SQSGGVTFQFDVSONNFMGTGKHYNASFSRSRSETREVYSLGWTNPYFTVNGVYSQSLSGYR 498
Db 430 GTESGISYQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTKDGV--SLGGNVF 487
Qy 499 KTKYDNK---NISNVYLDISYGSLSYGYPIDENQRISEGL-----NADNTKLHGREFMGI 550
Db 488 FENTDNKSQDTSNRYKRTTYGNTVTLGFPVNNNSYYVGLGHTYNTKISNFALEYNNRNLVI 547
Qy 551 SNVKOLMADGGKIQVDNNGIPDFKHDTYTNAILGNWYSSLDROPVFPQTQGSMSHYD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYFPFKGVKASLGRGRT 592
Qy 609 VFGDKTHQKVVYQGNIRPF-----IKKSVLRGKYAKLGYGN-NLPFFYENFYAGGYGS 660
Db 593 IPGSDNKYYKLSADVOGFYPLDRDLHWVWSAKASAGYAN-GFGNKRLPFYQTYTAGGIGS 651
Qy 661 VRGYDQSSLSRPSQAYLTARGQOTTLG-----EVVGGNALATFGSELILPLPFKGDWI 713
Db 652 LRGFAYGSIGPNA--IYAEYNGSGSGTFFKKISSDVIGGNAIATASAEELIVPTPFVSDK 708
Qy 714 I-DOVRPVIFIEGGQVFTT-GMDKQITDLTQFKDPOQATAEONAKAANRPLLTDQKQLRY 771
Db 709 SONTVRTSLFVDAASVWNTKWKSDKNGLESVDLK-----RLPDYGGKSSRIRA 755
Qy 772 SAGVGATWYTPIGPLSISYAKPLNKKQNDQTDTVQFQIGSVF 813
Db 756 STGVGFQWQSPIGLVFSYAKPIKKYENDVDEQFQFISIGGSF 797

RESULT 13
D71726
outer membrane protein omp1 (omp1) RPI60 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: D71726
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499
A:Accession: D71726
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-768 <AND>
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14627.1; PID:g386
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: omp1; RPI60
C:Superfamily: protective surface antigen D-15

Query Match 17.98; Score 754; DB 2; Length 768;
Best Local Similarity 26.38; Pred. No. 8.4e-36;
Matches 216; Conservative 145; Mismatches 383; Indels 76; Gaps 20;

Qy 7 KGFOVSAMTMAVMNTHAQAADFMANDITITGLQRTVTTIESLQSVLPFRLLGQVWSENQ 66
Db 2 KILSISKLTILLTIFYHIFADYVIRKTIIEGHNHRVERSTIESYLKLVNVTYNNKS 61

Qy 67 ADGVKALYATGNFSDVOVY-HQEGRIYQVTERPLIAEINFEGRNRLIPKEGLQGLKNAG 125
Db 62 DEALKRLYATSLFRINMYITNDGNLINVVTETTFPISVWFSGNSKITNLAKIETWMS 121

Qy 126 LAVGQPLKQATVQMIE TELTNQYISQGYNTTEITVKTMLDGNRVKLDMTFAECKPARV 185
Db 122 --GESLSQAELDKVKKILEIYKRSGRFSTKVTPKIKSLNNRVKVFIDIAEGPKTVIK 178

Qy 186 DINIIGNQHFSDADLIDVLAIKONK-INPLSKADRYTOEKLVTSLENRAKYLNAGFVR 244
Db 179 SIYFSGNEHYSDSELKSVLTFKESRWFLESDNIDPDVREYDKELREYQSVGFADF 238

Qy 245 EIKDAKLNINDEKNRIFVEISLHEGEQYRFQGTQFLGNLTQTAELALKFKAEEGFS 303
```

```
Db 239 RVISASVALNDKEVFTTYSIEBGEKYRFGNVTIDNKLNTINIKOLANKVINIKOGKIFN 298
Qy 304 QAMLEPQTNNISTKFGDDGYIAQRPVTRINDESRVTDVEYIDPVHPVYVRRINFTGN 363
Db 299 MKTVDDIAEKIGEYFANGYPAVNVYDIPKND-NHTADIKFIIEKADKVINKINIINN 357
Qy 364 FKTDQEVLRMRQLEGALASNQKIQLSRARLMRTGFFKHVTVDRPVPNSPDQVDVNFV 423
Db 358 LKTEDHVIRRAFKIEGDVMRSYIEKGERNLRLNDIEKVSISLAQT-KAKDKYDVNVE 416
Qy 424 VEEQPSGSTTAAGYSQSGVTFQDVSONNFMGTGKHVNASFSESRSEVYSLGTMNTPY 483
Db 417 VDEKSTSIGFDLGYNTAGLGRFSFLERNLVGTGKLLNAGVQVSKNSTSYGCGITDPH 476
Qy 484 FTVNGVSQSLSGY----RKTNDKNKNSVYLDVSGSLSGYGPIDE--NORISFGUNA 537
Db 477 FLDRDLSLSVNAFRNYTRGASVLTNTDQSYKLHSGIKVLSGYDMKEDLSHEIDYLKR 536
Qy 538 DNTKLHGGRFGMISNVKOLMADGGKIQVDNNGIPDKHDYTYTNAILGNWYSSSLDRPVP 597
Db 537 DILSAPSP-----SNSIFLNEQMGKLIITSAIG-----HTIT-----YDQTDNKKVP 577
Qy 598 TQG--MHSVVDLTVFGDKTHOKVYVQGYNIYRPFITKSVLRGYAKLG-----YGNLPLF 649
Db 578 KNGYLVSTQEFAGVGGDNKYIKHIECKFYKSFNNKVTLLKLSAAGDMAGLGKGMVRI 637
Qy 650 YENFYAGGYGVRGYDQSSLGPRSOAYLTARRGOQTTLGEVVGGNALATFGSELILPLPF 709
Db 638 SDRFNLDY-SIRFASGVGPR-----EKNTEGLGGRYTFTELNFPTPV 685
Qy 710 KGDWIDQVRPIFIEGGQVDTTGMKQTDIDLTFQDKPOATAEQNAKAANRPLLTQDKQL 769
Db 686 PEEF--NFTGAVFIDLGSVWG-VGLNKK-----QYKTPNG-----FYNDQSL 724
Qy 770 RYSAGVGATWTPIGPLSISYAKPLNKKQNDQTDVQFQI 809
Db 725 RASVGFGEIWTFRAPIRMDGWFPIKKQYD--DTQNEHL 762

RESULT 14
D97527
ompl protein precursor (U51683) [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: D97527
A.: Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: D97527
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <R>
A:Cross-references: GB:AE007869; PIDN:AAK87173.1; PID:g15156447; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2554
A:Map position: circular chromosome

Query Match 17.5%; Score 734; DB 2; Length 774;
Best Local Similarity 26.4%; Pred. No. 1.2e-34;
Matches 219; Conservative 150; Mismatches 354; Indels 106; Gaps 22;

Qy 15 TMAYVMVSTHAQAADFANADITITGLQRTVIESLQSVLPRLGQVYVSENLADGVKALY 74
Db 22 SVAGLGLVASAGVANAAYIKIDVRGAERSGADSVRSNITITAPKGNFNSDIDESVKRLY 81
Qy 75 ATGNFSDVQVYHQEGRIITYQVTRPLIAEINFEGNRLIPKEGLQEGLNAGLAVGQPL-- 132
Db 82 ATGYFSNVMSRVSGSTLVVTVNEQLNVQVVFNGNRKIKDD-----KLAGIVQTQMPG 135
Qy 133 -KQATVQMIELTNQYISQGYNYNTEITVTKQTMLDGNRVRKLDMTFAEGKPARVVVDINIIG 191
```

```
Db 136 FNOAVTADIAIKEAYSAGIRSDVEITTTQTVSVGGVRNIAFVINEGERTKIGRIDFIG 195
Qy 192 NOHFSDADLIDLVL-AIKONKINPLSKADRYTOEKLTVSLNLRKAYLNAGFVRFEIKDAK 250
Db 196 NNSYSDGLAALAVINTPKSNMLSFLTRKDYNEDKLRADDEALRQFYNNRGYADFRVSSD 255
Qy 251 LNIENDKNRIFVEISLHEGEQYRFQGTQFLGNLTVTO-AELEALLKFAAEFGSQAMLBO 309
Db 256 AVLDESKNEYIISITVDEGKKYDFGNVAVESTVPGVDSGSELQGLVETROGASYSKEVQO 315
Qy 310 TTNNTSKFGDDGYIAQIRPVTRINDESRVTDVEYIDPVHPVYVRRINFTGNFKTQDE 369
Db 316 SMEALSKRVAEGYFARVTPRGDRDMSGNTIGVTYIVDQGERAYVERIEIRGNTRTD 375
Qy 370 VLRRMRQLEGALASNQKIQLSRARLMRTGFFKHVTVDRPVPNSPDQVDVNFVVEEQPS 429
Db 376 VIRREFDISEGDAFNQTTITAAKRRLAALGYFSKYNISITAG-GSAPDRVYIVVDVEDQST 434
Qy 430 GSSTIAAGYSQSGVTFQDVSONNFMGTGKHVN-ASFSSRSETRVYSLGTMNTPFTVNG 488
Db 435 GSFGAGYSQNDGVLLEASVEEKNFLGRGQYIRVAAGAGEDDARTYSLSTPEYF---- 490
Qy 489 VSQSLSGYRYRTKYDNKNISNVYLDVSGSLSGYGPIDENQIRISFGLNADNTKLHGGRPM 548
Db 491 LGYRLAAGFDLTKNQSKESEDYNYDEQGFALRVTPATITENLSTTFKY----- 537
Qy 549 GISNVKQLMADGGKIQVDNNGIPDKHDYTYTNAIL---GW-----NYSSLDPRPVF 596
Db 538 ---TYKQINYEKGQDWQNNALAE-----PYQALIRGEDMTQOSILSTNLNTLDDRNM 588
Qy 597 PTQGMHSHVDLTVGF---GDKTHOKVYVQGYNIYRPFITK---SVLRGYA--KLGYGN 646
Db 589 PREGHQAA--LTNEFAGLGGDSEYKIKAKARYTYLSDYDVIGSLGQAGHVMTGDN 646
Qy 647 LPFYENFYAGGYGVRGYDQSSLGPRSOAYLTARRGOQTTLG-EYVGGNALTFGSELIL 705
Db 647 LLVFDQFRFGG-RQVRGFKNDGIGPR-----IGSDSIGGTTTYFAASAENVTA 691
Qy 706 PLP-FKGDWIDQVRPIFIEGGQVDTTGMKQTDIDLTFQDKPOATAEQNAKAANRPLLT 764
Db 692 PMPGPVPEDF--GLRLAGFVDAGTMYGNKVSQTV-----KD----- 726
Qy 765 QDKQLRYAGVGATWTPIGPLSISYAKPLNKKQNDQTDVQFQIGSVF 813
Db 727 -DNSIRASAGIGVMWASPFGPPIRVDAIPIAKEDYDEQORFPGMSNTF 774

RESULT 15
AE2746
group 1 outer membrane protein precursor ompl [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AE2746
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCI
: Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE2746
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-774 <R>
A:Cross-references: GB:AE008686; PIDN:AAL42387.1; PID:g17739796; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: ompl
A:Map position: circular chromosome
```

Query Match	17.58;	Score:734;	DB 2;	Length 774;
Best Local Similarity	26.4%;	Pred. No. 1.2e-34;		
Matches 219;	Conservative 150;	Mismatches 354;	Indels 106;	Gaps 22;
Qy	15	TMAVMYMNSTHAQAADFMANDITITGLQRVTIESLQSLVPLRGLGVVSENOLADGVKALY	74	
Db	22	SVAGLGVLASGVANAAVISKIDVRGBASRGADSVRSNITITAPKNTFSNDSIDESVKRLY	81	
Qy	75	ATGNFSDVVYHQBGRRIYQWTERPLIAEINFEGRNLIPKEGLEGKLNAGLAVGQPL--	132	
Db	82	ATGYFSNVSMRVSGSTLVVTVNENQLNVQVFNGRKTKDD-----KLAGIVQTQPMGP	135	
Qy	133	-KQATVQMIETELTNQISQGYNTETVTKQTMGLDGNRVLKMTFAEGKPARVVDINLIG	191	
Db	136	FNQAIVTADIARIKEAISAIGRSODEITQTQVSGQGVNATFVINEGERTKIGRIDFIG	195	
Qy	192	NQHFSDADLIDLVL-AIKDNKINPLSKADRYTQERKLVTSLENLRKAYLNAGVFRFEIDAK	250	
Db	196	NNSYSDGLAAAVINTKSNMLSFLTRKDVYNEDKLRADAEALRGALFYNNRGYADFRVYSSD	255	
Qy	251	LNIINEDKNRIPVETLSLHEGEOYRGQTFQLNLNLYTQ-AEELALKFKAEGFSQAMLEQ	309	
Db	256	AVLDESKNXYTISTVDGKGYDEGNVAVETSPGVGSELOGLIVETROGASYSKAKEVQQ	315	
Qy	310	TTNNISTKFGDGGYIYAQIRPVTRINDESRVDVEYYIDPVHVVVRRINFTGNFKTQDE	369	
Db	316	SMEALSKRVAGEYPPFARVTPRGDRDMSGNTIGVYIVDQGERAYVERIEIRGNTRTDY	375	
Qy	370	VLRREMRQEGALASNOKIQLSRARLMRTGFFKHVTVTRVPVNSPDQDVNFVVEEQPS	429	
Db	376	VIRREFISEGDAFNQTIITAARKRLAEGLYFSKVNISTAG-GSAPDRVIVVDVEDQST	434	
Qy	430	GSSTIAAGYSQSGVTFQFDYSONNFWCTGKHVN-ASFPSRSETREVSLSGNTNFTYVNG	488	
Db	435	GSFGIAGYSONDQGVLLAEASVEEKNFJLGRGOYIRVAGAGGEDDARTYSLSFTEPYF	490	
Qy	489	VSQSLSGYYRKTDKNKNISNVLDSYSGSLSGYPIDENQRIISFGLNADNTKLHGGRFM	548	
Db	491	LGYRLAAGFDLFKNQSKEDEYNYDEQGFALRVTPAITENLSTTFKY-----	537	
Qy	549	GISNVKQLMADGGKIVDNNIGIPDKHDYTYNNAIL-----GW-----NYSSLDRPVF	596	
Db	538	---TYKQINYEKGQWQNNALAE-----PYQALIRGEDWTQSTLSWTLNATLDDRNM	588	
Qy	597	PTQGMHSVDLTVGF---GDKTHQKVYVQNIYRPFIKK-----SVLRGYA--KLGYGNN	646	
Db	589	PREGQAA--LTNEFAGLGGDEYYKIYAKARYYYTSLDEYDVLGSLTGQAGHVMTGDN	646	
Qy	647	LPFTENFYAGYSGVRYDQSSLGRSQAYLTARRGQOTTIG-EVVGNAIATATGSELIL	705	
Db	647	LLVPDQFKFEG-RQVRFGKNDGIGPR-----ICSDIGSTGYTFAASAEEVTA	691	
Qy	706	PLP-FKGDWIDQVPRVPIFIEGGQVFDTTGMBKQITLDLTQFKDPQATQAKAANRPLLT	764	
Db	692	PMPGVPEDF--GLRLAGEVDTAGTWGNKVSTSQV-----KD-----	726	
Qy	765	QDKLRYASGAGVATWYTPIGPLSISYAKPLNKKQNDQTDTVQFGISYVF	813	
Db	727	-DNSIRASAGIGVMAWSPFGPIRDVYDAIPIAKEDYDEORFERFGMSNFT	774	

Search completed: September 5, 2002, 09:55:54
Job time: 273 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 09:54:47 : Search time 36.65 Seconds
(without alignments)
858.908 Million cell updates/sec

Title: US-09-701-711-2

Perfect score: 4202

Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDVTVOFGISVF 813

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1210	28.8	810	1	UP05_ECOLI
2	1053.5	25.1	795	1	D152_HAEIN
3	1052.5	25.0	793	1	D153_HAEIN
4	1052.5	25.0	797	1	D151_HAEIN
5	641	15.3	617	1	Y237_BUCAI
6	256.5	6.1	578	1	YTFM_HAEIN
7	227	5.4	577	1	YTFM_ECOLI
8	197.5	4.7	469	1	G551_HUMAN
9	171	4.1	463	1	U140_DROME
10	157.5	3.7	1577	1	HLXA_PROMI
11	134.5	3.7	1256	1	MRP_STRSU
12	149	3.5	1902	1	P3P_LACLC
13	148	3.5	475	1	YDF6_SCHPO
14	146	3.5	826	1	YEH8_ECOLI
15	146	3.5	1902	1	P1P_LACLC
16	143	3.4	1902	1	P2P_LACPA
17	142.5	3.4	705	1	PPCF_FLAME
18	140.5	3.3	1902	1	P2P_LACLC
19	140	3.3	2201	1	TENA_HUMAN
20	139.5	3.3	1324	1	CUT3_SCHPO
21	137.5	3.3	735	1	FCT_ERWCH
22	136	3.2	917	1	HXA3_HAEIN
23	135.5	3.2	1332	1	KXDO_BAGSU
24	134.5	3.2	866	1	YC8S_ECOLI
25	134.5	3.2	1178	1	CIAC_BACTK
26	134	3.2	655	1	YKDA_MYCCA
27	133.5	3.2	937	1	C532_ECOLI
28	133	3.2	398	1	YLB1_CAEEL
29	133	3.2	1048	1	AGOL_ARATH
30	132.5	3.2	863	1	YEJO_ECOLI
31	130.5	3.1	1957	1	YD86_SCHPO
32	128.5	3.1	905	1	HXA1_HAEIN
33	127.5	3.0	839	1	YDDB_HAEIN

34	127	3.0	484	1	YNC6_YEAST
35	127	3.0	922	1	DFO1_RICFE
36	126.5	3.0	1258	1	ICEN_ERWHE
37	126	3.0	925	1	VPH_BPHP1
38	126	3.0	1034	1	ICEN_PANAN
39	125.5	3.0	880	1	LYTD_BACSU
40	125.5	3.0	2022	1	ANTI_ONCVO
41	124.5	3.0	1341	1	VG37_BPT2
42	124.5	3.0	1754	1	PMPB_CHLFR
43	124.5	3.0	2366	1	TOXB_CLODI
44	124	3.0	1630	1	MSPL_PLAFK
45	124	3.0	1639	1	MSPL_PLAFW

ALIGNMENTS

RESULT 1
UP05_ECOLI
ID UP05_ECOLI STANDARD: PRT: 810 AA.
AC P39170; P39181; P77465;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Unknown protein from 2D-page spots M62/M63/O3/O9/T35 precursor.
GN YAEI OR B0177 OR Z0188 OR ECS0179.
OS Escherichia coli, and.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
RN [5]
RP SEQUENCE OF 21-32 AND 351-362.

P53969	saccharomyc
Q9raa9	rickettsia
P16239	erwinia her
P51735	bacterioph
Q47879	pantoea ana
P39848	bacillus su
P21249	onchocerca
P07067	bacterioph
O84418	chlamydia t
P18177	clostridium
P04932	plasmodium
P04933	plasmodium


```

Query Match      25.1%; Score 1053.5; DB 1; Length 795;
Best Local Similarity 31.8%; Pred. No. 1.6e-54;
Matches 260; Conservative 158; Mismatches 337; Indels 63; Gaps 20;

QY 23 STHAQAADFMANDITITGLQRTVIESLQVLPFRILGVVSENQADGVKALYATGNFSDV 82
DB 14 TTTVFAAPFVAKDIRVGVQDLEQIRASLPVRAQORTDNDVANIIVRSFLVSGRFDV 73

QY 83 QVYHOGRI-IQVTERPLIAIEINFEGRNLIPKEGLQGLKNAGLAVGLOPLKQATVOMIE 141
DB 74 KA-HOEGDVLVSVVAKSIISDKVIGNSIIPTEALKQNLNDANGFKVGVDLIREKNEFA 132

QY 142 TELTNOYISQGYNTETITVQTMLDGNRVKLDMTFAEGKPARVDINIIGNHFSADLI 201
DB 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASUTFFKGNESVSSLTQ 192

QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNRAKYLINAGVFRIKADAKLINTNEDKNRIF 261
DB 193 EQMELQPDSSWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251

QY 262 VESILHEGEQYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
DB 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRRSADIADVENAIKAKLGE 311

QY 321 DGYVYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRREMRQLEG 380
DB 312 RGYGNTTVNSVPDQDANKTLAITFVVDAGRRLTVQLRFEQNTVSADSTLQKQEMRQEG 371

QY 381 ALASNQIKQLSRARLMRTGFFKHVTVDR--PVPNSPDQVDVNFVVEQPSGSSTIAAGY 438
DB 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYKVKERTGSGINFGY 429

QY 439 SOSGGVTQFQVDSQNNFMGTGKHVNASFSRSETREVYSLGMTNPYFTVNGVSQSLSGYR 498
DB 430 GTEGISTQTSIKQDNFLGTGAASVIACTKNDYGTGSVNLGYTEPYFTKDGCV--SLGGINF 487

QY 499 KTKYDNK---NLSNVLDYSYSGSLSGYGPIDENQISFGL-----NADNTKLHGGRFMGI 550
DB 488 FENYDNSKSDTSSNYKRTYGSNTVLGFPVNNNSYVGLGHTYKINSFALEYNRNLYI 547

Query Match      25.0%; Score 1052.5; DB 1; Length 793;
Best Local Similarity 31.8%; Pred. No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 65; Gaps 21;

QY 23 STHAQAADFMANDITITGLQRTVIESLQVLPFRILGVVSENQADGVKALYATGNFSDV 82
DB 14 TTTVFAAPFVAKDIRVGVQDLEQIRASLPVRAQORTDNDVANIIVRSFLVSGRFDV 73

QY 83 QVYHOGRI-IQVTERPLIAIEINFEGRNLIPKEGLQGLKNAGLAVGLOPLKQATVOMIE 141
DB 74 KA-HOEGDVLVSVVAKSIISDKVIGNSIIPTEALKQNLNDANGFKVGVDLIREKNEFA 132

QY 142 TELTNOYISQGYNTETITVQTMLDGNRVKLDMTFAEGKPARVDINIIGNHFSADLI 201
DB 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASUTFFKGNESVSSLTQ 192

QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNRAKYLINAGVFRIKADAKLINTNEDKNRIF 261
DB 193 EQMELQPDSSWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251

QY 262 VESILHEGEQYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
DB 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRRSADIADVENAIKAKLGE 311

QY 321 DGYVYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRREMRQLEG 380
DB 312 RGYGNTTVNSVPDQDANKTLAITFVVDAGRRLTVQLRFEQNTVSADSTLQKQEMRQEG 371

QY 381 ALASNQIKQLSRARLMRTGFFKHVTVDR--PVPNSPDQVDVNFVVEQPSGSSTIAAGY 438
DB 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYKVKERTGSGINFGY 429

QY 439 SOSGGVTQFQVDSQNNFMGTGKHVNASFSRSETREVYSLGMTNPYFTVNGVSQSLSGYR 498
DB 430 GTEGISTQTSIKQDNFLGTGAASVIACTKNDYGTGSVNLGYTEPYFTKDGCV--SLGGINF 487

QY 499 KTKYDNK---NLSNVLDYSYSGSLSGYGPIDENQISFGL-----NADNTKLHGGRFMGI 550
DB 488 FENYDNSKSDTSSNYKRTYGSNTVLGFPVNNNSYVGLGHTYKINSFALEYNRNLYI 547

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
NCBI_TaxID=727;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN-PAK 12085;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
  Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
  species and may represent a universal protective antigen against
  invasive disease.";
RL Infect. Immun. 65:3701-3707(1997).
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY LOCATIONS TO THE SURFACE ANTIGEN D15 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
CC EMBL; U60834; AAB61977.1; -.
DR InterPro; IPR00184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surface_Ag; 1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 793 AA; 87511 MW; 51BFDB2036801A1A CRC64;

Query Match      25.0%; Score 1052.5; DB 1; Length 793;
Best Local Similarity 31.8%; Pred. No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 65; Gaps 21;

QY 23 STHAQAADFMANDITITGLQRTVIESLQVLPFRILGVVSENQADGVKALYATGNFSDV 82
DB 14 TTTVFAAPFVAKDIRVGVQDLEQIRASLPVRAQORTDNDVANIIVRSFLVSGRFDV 73

QY 83 QVYHOGRI-IQVTERPLIAIEINFEGRNLIPKEGLQGLKNAGLAVGLOPLKQATVOMIE 141
DB 74 KA-HOEGDVLVSVVAKSIISDKVIGNSIIPTEALKQNLNDANGFKVGVDLIREKNEFA 132

QY 142 TELTNOYISQGYNTETITVQTMLDGNRVKLDMTFAEGKPARVDINIIGNHFSADLI 201
DB 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAEILIQINEDDKAKLASUTFFKGNESVSSLTQ 192

QY 202 DVLAIKDNKINPLSKADRYTOEKLVTSLNRAKYLINAGVFRIKADAKLINTNEDKNRIF 261
DB 193 EQMELQPDSSWKL-WGNKFEQAQFEKDLQAIIRDYLLNNGYAKAQITKTDVQLNDEKTKVN 251

QY 262 VESILHEGEQYRFGOTQFLGNLTYYQAELEALLK-FKAEEGFSQAMLEQTTNNISTKFGD 320
DB 252 VTIDVNEGLQYDLRSARIIGNLGMSAELEPLLSALHLNDTFRRSADIADVENAIKAKLGE 311

QY 321 DGYVYAAQIRPVTRINDESRTVDVEYIDPVHPVYVRRINFTGNFTQDEVLRREMRQLEG 380
DB 312 RGYGNTTVNSVPDQDANKTLAITFVVDAGRRLTVQLRFEQNTVSADSTLQKQEMRQEG 371

QY 381 ALASNQIKQLSRARLMRTGFFKHVTVDR--PVPNSPDQVDVNFVVEQPSGSSTIAAGY 438
DB 372 TWYNSQLVELGKIRLDRGTGFFE--TVENRIDPINGSNDEVDVYKVKERTGSGINFGY 429

QY 439 SOSGGVTQFQVDSQNNFMGTGKHVNASFSRSETREVYSLGMTNPYFTVNGVSQSLSGYR 498
DB 430 GTEGISTQTSIKQDNFLGTGAASVIACTKNDYGTGSVNLGYTEPYFTKDGCV--SLGGINF 487

QY 499 KTKYDNK---NLSNVLDYSYSGSLSGYGPIDENQISFGL-----NADNTKLHGGRFMGI 550
DB 488 FENYDNSKSDTSSNYKRTYGSNTVLGFPVNNNSYVGLGHTYKINSFALEYNRNLYI 547
```

```

QY 551 SNVKQLMADGGKIQVDNNGIPDFKHDYTYNAILGNWYSSLDPRVFPFTQGMSSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYPTKGVKASLGGRTV 592
QY 609 VGFQDKTHQKV--YQG-----NIYRPFIKKSVLRGVAKLGYGN-NLPFFYENYAGGYGS 660
Db 593 IPGSDNKYYKLSADVQGFYPLDRDHRWWSAKASAGYAN-GFGNKRLLPFYQTYTAGGIGS 651
QY 661 VRGYDOSSLGPRQAYLTARRGOQT---TLGEVVGGNALATFGSELILPLPFKGDWI-DQ 716
Db 652 LRGFAYGISGPN-----IYAEHNGTFNKISSDVIGGNAITTAELIIVPFVSDKSQNT 708
QY 717 VRPVIFTEGGQVFDTT-GMDKQITDITQFKDPOQATAEQNAKAANRPLLTQDKQLRYSGAV 775
Db 709 VRTSLFYDAASVWNTWKSKDNGLSKVLKD-----LPDYGKSSRIRASTGV 755
QY 776 GATWYTPIGPLSISYAKPLANKKQNDQDTVQFQIGSVF 813
Db 756 GFQWQSPIGPLVFSYAKPIKKYENDDVQEQFQSIGGSF 797

RESULT 4
D151_HAEIN STANDARD; PRY 797 AA.
AC P46024;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (p-15-Ag) (Outer membrane protein p15).
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=SEROTYPE B;
RX MEDLINE=95255676; PubMed=7737523;
RA Flack F.S., Loosmore S., Chong P., Thomas W.R.;
RT "The sequencing of the 80-kDa D15 protective surface antigen of
RT Haemophilus influenzae."
RL Gene 156:97-99(1995).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=MINNA, AND BAGAN;
RX MEDLINE=97427952; PubMed=9284140;
RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
RA Klein M.H.;
RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
RT species and may represent a universal protective antigen against
RT invasive disease."
RL Infect. Immun. 65:3701-3707(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U13961; AAA85645.1; -
CC EMBL; U60832; AAB61974.1; -
CC EMBL; U60833; AAB61976.1; -
CC InterPro: IPR000184; Bac_surfAg_D15.
CC Pfam: PF01103; Bac_surface_Ag_1.
KW Antigen; Outer membrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
SQ SEQUENCE 797 AA; 87675 MW; 2F93DE538696AF1B CRC64;

```

```

Query Match 25.0%; Score 1052.5; DB 1; Length 797;
Best Local Similarity 31.6%; Pred. No. 1.9e-54;
Matches 260; Conservative 160; Mismatches 333; Indels 69; Gaps 20;

QY 23 STIAQAADFMANDITITGLORVITIESLOSVLPRFLQGVSENQADGVKALYATGNESDV 82
Db 14 TTTVFAAPFYAKDIRVDVQDLEQQIRASLIPVRGARTDNDVANIRSLFVSREDDV 73
QY 83 QVYHOEGRI-IYQVTERPLIAEINFEGRNLIPKLEGLKNAGLAVGQPLQKQATQVOMIE 141
Db 74 KA-HQEGDVLVSVVAKSIISDVKIKNSVIPTALKQNLNDANGFKVDVILIRKLENEFA 132
QY 142 TELTNOYISOGYNTTEITVKQTLMDNRVLDMTTFABGKPARVVDIINIQHNSDDALI 201
Db 133 KSVKEHYASVGRYNATVEPIVNTLPNNRAELIIOINEDDKAKLASLTFKGNESVSSLTQ 192
QY 202 DVLAIKDNKNPLSKADRYTQEKLVTSLENLRAKYLNAGFVRFETKDAKLINEDKNRIF 261
Db 193 EQMELQPDSSWKL-WGNKFEQAQFKDLOSIRDYLNNGYAKAQITKTDVQNLDEKTKVN 251
QY 262 VEISLHEGEQYRFQOTQFLGNLYTQAELEALLK-FKAEEGFSQAMLEQTTNNISTKPGD 320
Db 252 VTIDVNEGLOYDLRSARLIGNLGMSSAELEPLLSALHLNDFRSDIADVENAIKAKIGE 311
QY 321 DGYYIAQIRPVTRINDESRTVDVEYIDPVHPVTVRRINFTGNFKTQDEVLRMRQLEG 380
Db 312 RYGSATVNSVPDPDDANKTLAITLVVDAGRLRTVRLQRFEGTNSADSTLRQEMROQEG 371
QY 381 ALASNQIKLSRARLMRTGFKFKHTVDTR--PVPNSPDQVDVNFVVEBQPSGSSTIAAGY 438
Db 372 TWYNSQLVELGKIRLDRGTGFE--TVENRIDPINGSNDEVVDVYVKERTNGSINFGIGY 429
QY 439 SQSGGVTFQEDVSQNNFMGTGKHVNASFSRSETREVSLSGMTNPYFTVNGVSQSLSGYR 498
Db 430 GTEGSIYSQASVKQDNFLGTGAASVIACTKNDYGTSVNLGYTEPYFTKDGV--SLGDNVF 487
QY 499 KTKYDNK---NISNVYLDISYGSLSYGYPIDENQIRISGL-----NADNTKLHGGRENGI 550
Db 488 FENTDNKSQDTSNNYKRTYTSNVTLPFPVNNENSYVGLGHTYKNKISNFALEYNRNLYI 547
QY 551 SNVKQLMADGGKIQVDNNGIPDFKHDYTYNAILGNWYSSLDPRVFPFTQGMSSHVD--LT 608
Db 548 QSMK-----FKNGIKTNDFFS-----FGWYNLSNRGYPTKGVKASLGGRTV 592
QY 609 VGFQDKTHQKVYQGNITVYRP-----IKKSVLRGVAKLGYGN-NLPFFYENYAGGYGS 660
Db 593 IPGSDNKYYKLSADVQGFYPLDRDHRWWSAKASAGYAN-GFGNKRLLPFYQTYTAGGIGS 651
QY 661 VRGYDOSSLGPRQAYLTARRGOQTTLG-----EVVVGGNALATFGSELILPLPFKGDW 713
Db 652 LRGFAYGISGPN-----IYAEYNGSGTGTFFKISSDVIGGNAITTAELIIVTFVSDK 708
QY 714 I-DQVRPVIFIEGGQVFDTT-GMDKQITDITQFKDPOQATAEQNAKAANRPLLTQDKQLRY 771
Db 709 SQNTVRTSLFYDAASVWNTWKSKDNGLSKVLK-----RLPDYGKSSRIRA 755
QY 772 SAGVATWYTPIGPLSISYAKPLANKKQNDQDTVQFQIGSVF 813
Db 756 STGVGFQWQSPIGPLVFSYAKPIKKYENDDVQEQFQSIGGSF 797

RESULT 5
Y237_BUCAI STANDARD; PRY 617 AA.
AC P57331;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein BU237 precursor.
DE BU237.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
symbolic bacterium).

```

CC Bacteria; Proteobacteria; gamma subdivision; Buchnera.

```
OX NCBI_TaxID=1118099;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=Tokyo 1998;
MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
  Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -!- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
DR EMBL; AP001118; BAB12952.1;
DR InterPro; IPR000184; Bac_surfAg_D15.
DR Pfam; PF01103; Bac_surfaceAg_1.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 617 HYPOTHETICAL PROTEIN BU237.
SQ SEQUENCE 617 AA; 71807 MW; 9C91CA6D53BC27B8 CRC64;
```

Query Match 15.3%; Score 641; DB 1; Length 617;

Best Local Similarity 26.3%; Pred. No. 1.9e-30;

Matches 145; Conservative 132; Mismatches 261; Indels 14; Gaps 6;

Qy 16 MAVMVMS--THQAADFMANDITITGLQRTIESLSQSLVPLPRLGVVSENQADGVKAL 73

Db 8 IAFLMFFSILVHAE-NKFTVDDIQPKLNFQSNALKNIPPRIGTISQYDVKNRSKSL 66

Qy 74 YATGNFSDVQVYHOGRIYQVTERPLIAENFEGNRIPLKEGLEGLKNAGLAVGQPLK 133

Db 67 FRTGRFEDIKVSFLGKTIVNIRERPIISVSGNIIASSVLDPLKLNIEGTGKSFN 126

Qy 134 QATVOMIETELTNQISQYNTETVQOTMLDGNRVKLDWTFAGKPARVVDINIQN 193

Db 127 NFFTNLTKTIDFIDGRCKPDIKILKIFSKNNSVSIKILNIEGTGTEIKSILGVQ 186

Qy 194 HFSADLDLVLAIKD--NKINPLSKADRYTOEKLVTSLNLRKYLINAGVFREFIKDAKL 251

Db 187 AFSKEKILSLFKLDYHSHWNLLSKS-TYSPKELNDLEHLKNFVLSNGYFFNVTNKKV 245

Qy 252 NINEDKNRIFVEISLHEGEQYRFQOTFLGNLTYYTQAELEALLKPKAEFGSQAMLEQTT 311

Db 246 DFLQDKKQVDITIGVSEGGKYKISNFFVNGNLPFYQKLITLININPNEFYNRDKIDIIV 305

Qy 312 NNISTKFGDGGYVYAIIRPVTRINDESRVDVEYVIDPVHPVYVRRINFTGNFKTQDEVL 371

Db 306 NKITFLSEYGVNTKVIQVDPOIDHKKKTIALNFNIDMKKRYFVKRIYFTGNEITQDRLV 365

Qy 372 REMQLSGALASNOKIOLSRARLMRTGTFKHVTVDTVPVPSPDQVDVNFVVEEQSGS 431

Db 356 RKKIKOMEGKFTNKLVELGKLLKTKYFSDVKIIOQLNSYDSNQIDITYQVKEQTGS 425

Qy 432 STIAAGYSQSGVTFQFDVFSQNNFMGTGKHVNASFSSRSTREYVSLGMTNPNFTVNGVSQ 491

Db 426 INFGLGYGDSCTSFNLAFSQDNIFGSGNSLKVDDIKNQYKLDISTISYPFFYNADL 485

Qy 492 SLUGYRTTKYDNKNISNVLDYSYGSLSYGYPIDENQRISFG-----LNADNTKLHG 544

Db 486 NARFFYNDFKNFYDNISIKTKTYFEGNLGLPLNNYKNVNFYGTHTNSINNEEKKIH- 544

Qy 545 GFMGISNVKQL 556

Db 545 GKFTGELNINKL 556

```
RESULT 6
YTFM_HAEIN
ID YTFM_HAEIN STANDARD; PRT; 578 AA.
AC P44038;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein HI0698 precursor.
GN HI0698
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
  Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
  McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
  Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
  Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
  Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
  Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
  Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
  Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
  influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
  Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC -!- SIMILARITY: STRONG, TO E.COLI YTFM.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC
CC EMBL; U32752; AAC22357.1;
DR TIGR; HI0698;
KW Signal; Complete proteome.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 578 PROTEIN HI0698.
SQ SEQUENCE 578 AA; 65812 MW; 34F9ACL89C505876 CRC64;
```

Query Match

Best Local Similarity 6.1%; Score 256.5; DB 1; Length 578;

Matches 132; Conservative 94; Mismatches 219; Indels 267; Gaps 25;

Qy 152 GYNTETITVKQPMLOGNRVKLDWTFAGKPARV--VDINIGNQHFSDADLDVLAIKDN 209

Db 76 GYESSVPERQKQCKRDLIAHVTPGPTKIAGTDVQIEG-----AAQDE 123

Qy 210 KINPLSK-----ADRYTOEKLVTSLNLRKYLINAGVFREFIKDAKLNINEDNRIF 261

Db 124 NFNALRKNLPKQDVLVLEHOTYDDYKTAISRLA---LNRGYFDGNFKISLEISPTHQAW 180

Qy 262 VETSLHEGEQYRFQOTFLGNLTYYTQAE-----LEALLKFAEEGFSQAMLEQTTNIST 316

Db 181 WRMLFDSGVRYHY-----GNITFSSHQTRDDYLNILNIKSGDPY-----LMNLSL 227

Qy 317 KFGD--DGYYYAQIRPVTRINDESRVDVEYVIDPVHPVYVRRINFTGNFKTQDEVLRRE 374

```
Db 228 LTSDPSSNPFSSVLQPNVNHKSKTVDEILYP-----RK 264
Qy 375 MRLEGALASNOIKLSRRLMRTGFFKHVTVDRPVPNSPDQVDVNFVEEQSGSTI 434
Db 265 KNAME-----L 270
Qy 435 AAGYSQSGVTFQFDVSNQNMGTGKHVNASFSRSETREVYSLGNTNPFYTVNGVS-----490
Db 271 VGFTDGGVHGQ-----IGWTFWINSRSHLSRN 301
Qy 491 -----QSLSGYRKTKYDNKNISNVYDSYSGSLSYGYPIDENQRISFGLNADNTKL- 542
Db 302 LYLAPKOTLEATYRMLPL--KNPLNYID-----FAVGWGEKE-----NDTNRVL 347
Qy 543 -----HGRFPMGINSVQKLMADGGKIQVDNNGCIPDKHDTYTNAILGNWYSSL 591
Db 348 TLSALRYWNNAHGWYFG-----GLRMRYDSFTQADITDKTLLLYPTVGTFTTRL 397
Qy 592 DRPVFTQG-----MSHSDVLTGFGDKTHQKVYQGNIRYRPFIRKSVLRGYAKLGYGN-- 645
Db 398 RGGSFATGWDYOKITFDLSKRIWLSSEFFIKVQASSAWVRVYAENHRVVARAEIGYLHTK 457
Qy 646 ---NLPPFENYAGYSGVGRYDOSSLGPRSQAYLTARRGOQTTLGEVVGGNALATGSE 702
Db 458 GIEKIPPTLRFAGGDRSVRGYGYKKIAPKNRN-----GKLVGGSRLITTSLE 505
Qy 703 LILPLPFKGDWDQVRP---VIEFGQGVFTDTGMKDQITDLTQFKDPQATAEONAKAA 758
Db 506 ---YQYQYPNWAATFADSG-----LAADNYTA- 531
Qy 759 NRPLLTDQKRLYSAGYAGWTYPTIGPLSISYAKPLNKKQNDQTDTVQFQIG 810
Db 532 -----KELRYGTGCVGRNAPVGAIKFDIATPIRDKNSK--NIQFYIG 573

RESULT 7
YTFM_ECOLI
ID YTFM_ECOLI STANDARD: PRT; 577 AA.
AC P39320;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ytfm precursor.
GN YTFM OR B4220 OR Z5831 OR ECS198.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RL Nucleic Acids Res. 23:2105-2119(1995).";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).";
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-O157:H7 / RIMD 0509952;
```

```

MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
CC -I- SIMILARITY: STRONG, TO H. INFLUENZAE HI0698.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U14003; AAA97116.1; -;
DR EMBL; AE000493; AAC77177.1; -;
DR EMBL; AE005654; AAG59418.1; -;
DR EMBL; AP002568; BAB38621.1; -;
DR EcoGene; EG12513; ytfm
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 577 HYPOTHETICAL PROTEIN YTFM.
SQ SEQUENCE 577 AA; 64796 MW; 253D1D5BEA744D25 CRC64;

Query Match 5.4%; Score 227; DB 1; Length 577;
Best Local Similarity 20.5%; Pred. No. 3.5e-06;
Matches 151; Conservative 82; Mismatches 238; Indels 264; Gaps 31;

Qy 115 EGOBLNAGLAVGOPLKQATVQMIETELTNQYISQGY--NTEITVKQTMLDGNRVKL 172
Db 63 DAIREGLK-----ALGYQYQTFIEDLRPPKGRQV-L 94
Qy 173 DMTFAEGKPARV--VDINIIGNQHFSDADLIDVLAIKONKINPLSKADRYTOEKLVTSL 230
Db 95 IAKVTPGVPLVIGGTDVVLRGAR-TDRDYKLKLLDTPAIGTVLNOGYENFKSLTSTA 153
Qy 231 NLRAKYLNAGVRFRIKDAKLINEDKNRIFVEISLHGEQYRFQGTQFLGNLTQVQABL 290
Db 154 -LRKYFDSEFTK-----AQLGIALGLHKAFWIDYNGSERVRFHGVFEG-----198
Qy 291 EALLKFAKEEGFSQAMLEQTNNISTKTFDDGYYAQIRPVPTRINDESTRVDEYVIDPV 350
Db 199 -----SQIRDE-----YLQNL 209
Qy 351 HPVYVRINFNGFKTQDEVLREMRQLEGALASNQIKLSRRLMRTGFFKHVTV---406
Db 210 VP-----FKGDEYESKDLAEL-----NRLSATGFWNSVVPAPQF 245
Qy 407 ----DTRPVNSPDQVDVNFVEEQSGSSTIAAGYSQSGG--VTFQFDVSNQNMGTGK 460
Db 246 DKARETKVLP-----LTGVVSPRTETIETGVGYSTVDGPRVKATKKPMWNSYGHSL 298
Qy 461 HYNASFSRSETREVYSLGM---TNP---YFTVNGVSQSLSGYRKTKYDNKNISNVYSSL 514
Db 299 TTSTISAPTEQTLDFSYKMPLLKPNLEQYLVQV-----GFKRTDLNDESDSTTLVAS 352
Qy 515 YGGSLSYGYPIDENQRISF-----GLNADNTKLHGGFRMGISNVKQLM 557
Db 353 RYWDLSUGGQWRALNRLWSDHDTQGEITNTMLTFPGVMISRTSRGG-----LM 402
Qy 558 ADGGKIQVDNNGCIPDKHDTYTNAILGNWYSSLDRPVEPTQGMHSHVSLDTVGFGDKTHQ 617
Db 403 PTWGDQ-----RYSIDYSN---TAWG-SDVDFSVFQAQN-----433
Qy 618 KVVYQGNIV--RPFTKSVLRGYAKLGYGNLPPFTENFYAGGYSVGRYDOSSLGPRSQ 675
Db 434 --WIRTLYDRHRFVTRGTL-GWIEGTGDFKVPDPLRFFAGGDRSIRGYKYKSIAPK---487
```


CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>) or send an email to license@isb-sib.ch.

DR	EMBL; M30186; AAA25657.1; -.
DR	PIR; A35140; A35140.
KW	Hemolysis; Toxin; Outer membrane; Signal.
FT	SIGNAL 1 29
FT	CHAIN 30 1577
SQ	SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;

Query Match	3.7%	Score 157.5;	DB 1;	Length 1577;
Best Local Similarity	19.1%	Pred. No. 0.18;		
Matches 200;	Conservative 133;	Mismatches 349;	Indels 365;	Gaps 50;

Qy	10	QVSAMTAVMMVWSTHAQAADFMANDIITIGLQRVITIESLQSVLPRL-----	57
Db	206	KITAAEISAFQTGNTSQHFIDILSSQKPSALDSYFFGSMQSG-RIRIINTABGSGVKLA	264
Qy	58	GOVVSENL---ADGVKALYATGNFSDVQV---YHQEGRIIYQVTERPLIAEINFEGR	110
Db	265	GKPTAONDSVKADNIQ---TDSQVRVDSYDKDGSENTQ---NYRGGI	306
Qy	111	LIPKEGLEQL-----KNAGLAV-----GOPLKQATV	137
Db	307	TVNNSSSQTLTKTELKGNKITLVASSHNOIKASDLMGDDITLQAGADLTIDGKLOQK	364
Qy	138	QMTELTNQYISQGYNTEITVKQTMLDGNRV--KLDMTFABGKPARVVD---INIIG	192
Db	365	---ETIDINRFYSWKYDVTKEKEIQOIGSGIDAKNNATLTATKGDVTDIAAKIN-AGN	420
Qy	193	QHFSDADLDVLAIKONKINPL---SKADRYTOEKLVTSLN-----LRAKYL	237
Db	421	N-----LAINAKNDIHNGLYEKESRESNGNKRNHTSLESGWSNSHQETELKASEL	473
Qy	238	NAGFVRPEIKD-----AKLTINE-----DK-----	257
Db	474	TAG---KDLGLDAQGSITAGQAKLHANENVLVNAKDININLVQKTNNDKTVTVDNHVM	527
Qy	258	-----NRIFVEISLHGEYOYRGOTOFGLN	282
Db	528	WGGIGGQNNNNQOVSHATOLTADGQLLLAADNNVNITGSQVKGNGQAFVKT--QGD	586
Qy	283	LTVTOAELEALLKFAEBGFSQAMLPOTNNISTK-----	317
Db	587	VVIDNALSETISKIDERTGTAFTNITKSSHNETNKQTSGSELSIDAQLTVVSGNDVNVI	646
Qy	318	-----FGDGGYVYAAQIRPVIRINDE-----SRVVDVEYIIDPVHPVY	354
Db	647	GSLIKSAADKLHLSGLDINVKSQAQ--QVTKIDDEKYSLAITHAKEVEDQKYSAGFH---	701
Qy	355	VRRINFTGNFKTODEVLRREMQLEGA---LASNOKIQLSRRLMTGFFKHVTVDTDRPV	411
Db	702	---ITHTTNKTSTET-EQANSTISGANVDLQANKDVTTFAGSDLKTTACNASITGD---	753
Qy	412	PNSPDQDVNFVVEEQ-----PSCGSITIAAGYSOSGV-----TQOFDVSQNNFMGTOK	460
Db	754	-----NVAFVSTENKKQTDNTDTITISGFSYTGVDKVGSKADFYD-----KO	797
Qy	461	HVNASFSPR---STREVIYSLGMT-NPVFTVNGVYSQSLSGYFKTKYDNKNISVYLD---	513
Db	798	HTQTEVTKNRGSQTEVAGDLTITANKDLLHEGASHHVEGRYQES---GENIQHLAVNDSE	854
Qy	514	-SYGSSLSYGYPIDENORISFGLNADNTKLHGRFRMGISN-VKQLMADGGKIQVDNNGIP	571
Db	855	TSKTDLSLVNG-----IDVGVNLD-----YSGVTKPKKKALEDGVTNTKPGNN-T	897
Qy	572	DFKHDVTTYNAILGW-NYSSLDPRVPFTQGMSHSVDLTVGFCDKTHQKVYVQGNIIYRPI	630
Db	898	DLTKKVTARDAIANLANLSNLETFP-----NVGVEVIGKGGSGQ-----QSQTDSQAV	944
Qy	631	KKSVLRGYAKLGYGNLPPYENFYAGYGSVGRGYDQSSILGPRSQAYLTARRGOQTTLGEV	690

[illegible]

RESULT	11		
MRP_STRSU			
ID	MRP_STRSU	STANDARD;	PRT; 1256 AA.
AC	P32653;		
DT	01-OCT-1993 (Rel. 27, Created)		
DT	01-OCT-1993 (Rel. 27, Last sequence update)		
DT	01-OCT-1994 (Rel. 30, Last annotation update)		
DE	Muramidase-released protein precursor (136 kDa surface protein).		
DE	GN		
OS	Streptococcus suis.		
OS	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Streptococcus.		
OC	NCBI_TaxID=1307;		
RP	[1]		
RN	SEQUENCE FROM N.A.		
RC	STRAIN=SEROTYPE 2 / D282;		
RC	MEDLINE=92267630; PubMed=1587602;		
RA	Smith H.E., Vecht U., Gielkens A.L., Smits M.A.;		
RT	"Cloning and nucleotide sequence of the gene encoding the		
RT	136-kilodalton surface protein (muramidase-released protein) of		
RT	Streptococcus suis type 2";		
RL	Streptococcus suis type 2";		
RL	Infect. Immun. 60:2361-2367(1992).		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.		
CC	-1- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS		
CC	IN THE REGION OF THE MEMBRANE ANCHOR.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X64450; CAA45781.1; -		
DR	PIR; A43829; A43829.		
DR	InterPro; IPRO01899; Gram_pos_anchor.		
DR	Pfam; PF00746; Gram_pos_anchor; 1.		
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; 1.		
KW	Repeat; Transmembrane; Cell wall; Signal.		
FT	SIGNAL	1	47
FT	CHAIN	48	1256
FT	DOMAIN	48	1231
FT	TRANSMEM	1232	1248
FT	DOMAIN	1249	1256
FT	REPEAT	663	681
FT	REPEAT	839	861
FT	REPEAT	953	1006
FT	REPEAT	1064	1084
FT	REPEAT	1089	1142
FT	REPEAT	1143	1195
FT	DOMAIN	1223	1228
FT	SEQUENCE	1256 AA; 135795 MW; DCF7F65242F14341	CRC64;
FT	QC		

Query Match

3.78; Score 154.5; DB 1; Length 1256;

Best Local Similarity 19.0%; Pred. No. 0.19;		Matches 199; Conservative 125; Mismatches 319; Indels 405; Gaps 55;	
QY	9	FQVSAMTMAVMVMSTHAQAADFMANDITITGLQVTVIESLQSVLPRLGQVSENOI--	66
Db	25	FGAASVLGLSVLGAQAQV---VKADETVAS-SEPTIAS--SAPASTEAVEAEAEKTN	78
QY	67	ADGKALYATGNFSDVQVYHOGELIYQVT-ERPLIAEINFEGNRLIPK---EGLOEGLK	122
Db	79	AENTSAVATTS--TEVE---KAKAVLEQVTSSEPLLACL---GOKELAKTEDATLAKAIE	130
QY	123	NA--GLAVGOPL---KQATVQMIETELT-----NOYISQGYNTEI-----	158
Db	131	DAQFLAAKAILADSEATVEQVAQAQVAAKVAEALGNELQKVTVDGLLLTAALDTVPD	190
QY	159	TVKOTML--DGNRVKLDMT-----FAGKPARVVDINIIGNHFSDADLDIVLAIKDNK	210
Db	191	TTASTLKVGDEGTLIDSTTATPSMAEPNGAAIAP-----HTLRTQDG- 234	
QY	211	INPLSKADRYTOEK--LVTSLNLRK-YLNA--GFVRFEIKD-----AKLNINED	256
Db	235	IKATSEPNWYTFESVDLSYKNKNASSYKGAEDVAYIRYSLDNDSSSTAVLAEL-VSRT	293
QY	257	KNRIFVEISLHGEQYRFGQFL----GNLTYT-----QAELEALKFKAEFGFSQAM	306
Db	294	TGDLVLEKTYTIEPGESVTFSPHTKVNANNSNITVTYDTSLSANTPGALKFSANDDVYSTI	353
QY	307	L-----EQTNNISLTKFGD-----DGYYY-----	325
Db	354	IVPAYQINTRYTIESGKVLATYGLQTIAGQVVPSSVRFTGYDYVATTKAVOGPPPK	413
QY	326	-----AQIRPVTRINDESRVDVEYIIDPVHVVVRRINFTGNFKTQDEVLRL	372
Db	414	GTVYLACTVQKDTQYKVIKREIVENDQAVLKPYLD---PIYKGEVDWRTG-----	461
QY	373	REMROLEGALASNOKIQLSRARLMRTGFKHKKVVDTRVPVNSPDQVDVNFVVEEPSSGS	432
Db	462	-----DTGTGTELLT-----TSPTTYKVGTYD-----	484
QY	433	TIAAGYSOGGVTFOFDV-----SONNFMGTGKHVNASFSRSETR-----EVV-	475
Db	485	-----YNINSKITADPTIDPTKNVWVFKESEQNEGSKYRVIAQNSGDETKGIVGKIYI	539
QY	476	-----SLGMTNPYFTVNGVSQSLSG-----YVRK 499	
Db	540	ATQVWTKLGTNKGWEDYSDDOAGIKFNKGFAGVQNTLRNATPATAVETTVYIKES	599
QY	500	TKY-----DNKNISNYLDSYGSLSYGYPIDENQRI-SFGLNADNT-----	540
Db	600	SKYGDVIVEYTDGKQIVNSVVDTPKSAIGTEYNTDVRPASPASLVAADGTVYFYKEVKS	659
QY	541	---KLHGGFRFGISNVKQLMADGGKIQ-----VDNNG-----IPDFKHDTYTTNAILGWN	588
Db	660	DSAKTGTGTVAGTTVTKYVYKAGSVNVNFVDINGKVIKAPVSDEK-----DAKPGYNY	713
QY	589	SSLDPRVPFTQGMHSHVDLTGFGDKTHQKVVYQGNYPFFTKKSVLRG---YAKLYGN	645
Db	714	DT-----DL-----DOKLASITFECKEY-----KLVPAGDYPVGVKVGKN	748
QY	646	NLPFFENFYAGG-----YGSVRG-----YDQS-----SL 669	
Db	749	NLIEVGNNTAKGIDPTTGKIEAGVNKEVTYVYRAVTVGSVVVNVKDFEGNVKIDPETDVS	808
QY	670	GPRSOAYLT-----ARRGQOTLT-----GEVVGGNALATFCSELILPLPPK	710
Db	809	APVGDAYTTDKKPNELIKGDSRYVLVPSKTDGEENGKVIETITVTYVYQKV-----	862
QY	711	GOWIDQVRPVIFIEGQV---EDTTGMDKQTDLDIQFKDQPAEQNAKAAANRPLLTQDK	767
Db	863	ANWIEIPNPETDRPKVPYDPEDTPEP-EPIDPT-----TPGTNGEVPNIP-----	908
QY	768	QLRYSAGVGATWYTPIGPLSISVAKPLN 795	

Db 909 ---YVPG-----YTPVDPKONTPLKPID 928

RESULT 12
P3P_LACLC
ID P3P_LACLC STANDARD; PRT; 1902 AA.
AC P15292;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE P111-type proteinase precursor (EC 3.4.21.96) (Lactocep
DE associated serine proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OG Plasmid.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 188-197.
RC STRAIN=SK11;
RC MEDLINE=89340435; PubMed=2760036;
RA Vos P., Simons G., Siezen R.J., de Vos W.M.;
RT *Primary structure and organization of the gene for a procaryotic,
RT cell envelope-located serine proteinase.;
RL J. Biol. Chem. 264:13579-13585(1989).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
CC specificity, although some subsite preference have been noted,
CC e.g. large hydrophobic residues in the P1 and P4 positions, and
CC pro in the P2 position. Best known for its action on caseins,
CC although it has been shown to hydrolyse hemoglobin and oxidized
CC insulin B-chain.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04962; AAA03533.1; ALT_SEQ.
DR PIR; A32634; A32634.
DR HSSP; P00782; 2SPT.
DR MEROPS; S08.019; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
KW SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876
FT TRANSMEM 1877 1895
FT DOMAIN 1896 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT DOMAIN 1867 1872
P111-TYPE PROTEINASE.
EXTRACELLULAR (POTENTIAL).
MEMBRANE ANCHOR (POTENTIAL).
CYTOPLASMIC (POTENTIAL).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CHARGE RELAY SYSTEM (BY SIMILARITY).
CONSERVED IN GRAM-POSITIVE COCCI SURFACE


```
FT SQ SEQUENCE 1902 AA; 200550 MW; 87CECBAA9345F9D3 CRC64;
Query Match 3.5%; Score 149; DB 1; Length 1902;
Best Local Similarity 19.6%; Pred. No. 0.74;
Matches 175; Conservative 116; Mismatches 356; Indels 246; Gaps 44;
Qy 4 SYFKGFQVSMATMVMV-MSTHAQAADFMAANDIIT---GLQVTTIESLQSVLPFRIG 58
Db 646 AYYKQKGTALDFLTKTVMNTAQIPINDINNNVIVSPRQAGLVYDKAAIDALEKPS 705
Qy 59 QVSNQIADGKALYATGNSDVQYHQ--EGRIYQVTERPLIAEINFENRLI----- 112
Db 706 TVVAEN---GYPAVELKDFSTDKTKFLTFTNRTTHTLTYQ---MDSNTDINAVYTSAT 758
Qy 113 -PKEGL--OEGLNAGLAVGQPLKQATVQMIELET-----NOYISQGYNTEIIVK 161
Db 759 DPNQGLVDKIDGAAIKAGSNITVPAGTQAEFTSLPKSFDOQQVVEGFLNFKGS-- 816
Qy 162 QTMLDGNRVKIDMT--FAEGKPARVVDINIIGNQHFSDADLDIVLAIKDNKINPLSKADR 219
Db 817 ---DGRNLNLYMFFGWDGKIVD-SLNGITYSPAGNFGTVPPLKKN---NTGTQ 867
Qy 220 YTOEKLVTSLNLRKAYLNAGFVRPEIKDAKLNINEDKNRIFVEISLH----- 267
Db 868 Y-YGGMVTDADGNKT-----VDDQAIASFSDSKNLYNDISMKYVYLLRNISNVQV 915
Qy 268 -----EGQYRFQGTQFLNLTQAELEALLKFAESGEFQSOAMLEQTTNNISKFGDGG 322
Db 916 DILDQGNKVTLSSTNRKTKTYNAHSQQYIYNA-PAWDGTYDYDQDGNIKT--ADGG 972
Qy 323 YYAQIRPVTRINDESRVDVEYYIDPVHPVVVRINF---TGNPKTQDEVLRRMRQLE 379
Db 973 SYTVRISGVEGGDKRQVDFPFKLDKAPT-VRHVALSAKTENGKTQYLYTAEAKDDL 1031
Qy 380 GALASNQIQLSRARLMTGTFKKHVTVDTRVPNSPDQVDNVFVVEQPSGSSITAAQYS 439
Db 1032 GLDAT-----KSVKTEINEVTN---LDATF-----TDAGTTADGTYT 1064
Qy 440 QSGGVTFQFQVSONFMCTGKHV-----NASFSRSETREVYSLGTMNPFYTVNG--- 488
Db 1065 K---IETPLSDEQAQALNGDNSAELYLTDNASNATDQASVQKPGSTFDLIVNGGGIP 1121
Qy 489 --VSQSLSGYRKTK-----YDNKNISNYVLDSYGGSLSYGYP 524
Db 1122 DKISSTTGYEANTQGGCTYTFSGTYPAADVGTYYTDAQGKKHDLNTTYDAATNFTASMP 1181
Qy 525 I-DENORISFGLNADNTKLHGRFMGINSVKOLMADGKGIQVDNNGIPDFKH----- 575
Db 1182 VTNADYAAQVDLYAD--KAHTQLLKHEFTKVLMAPTFTDLKFNNGSDQTSATIKVTGT 1239
Qy 576 ---DYTVYNAILGMNYSLSLRPVPTQGMSH-SVDLTGVFGDK-----THQ 617
Db 1240 VSADTKTVN--VHTVAALD-----AQHHSVDVPVNYGDNITKVATDKDGNITTEQ 1290
Qy 618 KVVYQGNIRP-FIKKSVLRGYAKLGYGNLPPFENFYAGGYGVRGVDQSSLSGPRSQAY 676
Db 1291 KTISS--YDPMLKSVT-----PDQGVKFGTNKE--NATSAKFYDPKT---GIAT 1335
Qy 677 LTARRQOTTIGEYVVGNA-----LATFGSELILPLPF-----KGDW----- 713
Db 1336 ITGKVKHPTTTLQVDGKQIPKIDDLTFSFTLDGLTGQK-----PFGVVVGDTQNKTFQ 1390
Qy 714 -----IDQVRPVIFIEGGQVDFDTGMDKQITDLTQFKDPQ-----ATAEQNAK 756
Db 1391 EALSFILDAVAP-----TLSDSSTDAPVYTNPNFQITGTADNAQ 1432
RESULT 13
YDF6_SCHPO STANDARD; PRT; 475 AA.
ID YDF6_SCHPO
AC Q10478;
```

```
YEHF_ECOLI
ID YEHF_ECOLI STANDARD; PRT; 826 AA.
AC P33341;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE Hypothetical outer membrane usher protein yehB precursor.
GN YEHF OR B2109.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
[3]
RP SEQUENCE OF 307-826 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
Kasai H., Kimura S., Kitagawa M., Kitagawa M., Makino K., Miki T.,
Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
Yamamoto Y., Horiuchi T.;
RA "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392 (1996).
CC -1- FUNCTION: INVOLVED IN THE EXPORT AND ASSEMBLY OF A FIMBRIAL
SUBUNIT ACROSS THE OUTER MEMBRANE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
(bY similarity).
CC -1- SIMILARITY: BELONGS TO THE FIMBRIAL EXPORT USHER FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
Matches 148; Conservative 92; Mismatches 306; Indels 176; Gaps 34;
QY 135 ATQMIEET-LTNOYISQGYNT-----ETVK-----QTMIDGNRVKL-----DMTAEAG 179
DB 42 ANRLDDNQPLPGYDIDIVNKNQWKGKYEIIVKDNQETCLSRVIRKLGINSDFNSG 101
QY 180 KPARVVDINIGNOHFADLIDVLAITKDNKINPLSKADRYTQEKLVTSLENLAKYLNA 239
DB 102 KQCLTFEQLVGGSGYTWDIGVF-----RDLDFSPQAAWVEEL 139
QY 240 GFYRFEIKDAKLINEDKNRIFVEISLHEGEYOYRFGOTQFLGNTLYTQAEALLKFAE 299
DB 140 GYVPE-----NWERGINAFYTSYL---SQY-YSYKASGNKNSYVRNSGLNLGW 189
QY 300 EGSQAMLEOTNTN-----ISTKFGDDGYAQAIRPVRINDESRTVDVEYIDVHPVY 354
DB 190 QLHSDASFSTNNPNPGVWKSNTLYLERG--FAQLLGLTRVGDWYSSDI---FDSVRF 244
QY 355 VRRINFTGNFKTODEVLRREMRQL-----EGALASNQKIQLSRRLMTGFFKH 403
DB 245 VR-----LFRDMQLPNSKQNETPRVQGIASNALVTIE-----QNGF--- 282
QY 404 VTVDTRVPNSPDV-----DYNFVVEEQPSGSSTIAAGYSQSGGVTFQFDYSON 453
DB 283 -VYQKEVPPGPFATLDLAGGADLDVSKADGSGVTIYLVYAAVPMN-LQPGVSKY 340
QY 454 NFMGTGKHVNASFSE--TREYVSLGTMTPYFTVNGVSQSLSGYRRTKYDNKNISNYVL 512
DB 341 DLAAGRSHIEGASKQSDVQAGYQYGFNN-LTLLYGGSWANNYYAFT----- 387
QY 513 DSYGGSISGYPIDENQRIEFLNADNTKLHGRFMGINSVKQLMADGKIQVNDNGLPD 572
DB 388 -----LGAGW-----NTRIG-AISVDATKSHSKDNGD-----VFDGQSYQIAYN--K 427
QY 573 FKHDYTYNALGLNYSLSLDRPVPPTQGMSSHVDLTVCFGDKTHQKVY----- 621
DB 428 FVSTSTFRFGLAARWYSSRDYRTFNDHWANKD-----NVRRENDVYDIADYQND 482
QY 622 QGNIYRPFIRKSVLRGAKLYGNLPPYENFYAGGYSVGRYDQS-SLGRPSQAY-LTA 679
DB 483 RKNFSANMSQSLPEGWGSV-----SLSTLWRDYWGRSGSSKDYQLSYNNLRISYTLAA 538
QY 580 RRGQOTILGEVVGGNALATFQSEILLPLPKGDWIDQV---RPVIFIEGQVEDTTGMDK 736
DB 539 SQAYDENHHEKREN-----IFISIPF--DWGDDVSTPRQIYMSNSTTFDDQGFAS 588
QY 737 QTIDLTQFKDQATAEQNAKAAARPLLTQDKOLRYSAGVGATWTPIGPLSISYAKPLNK 796
DB 589 NN---TGLSGTVGSRDQFNQGVN--LSHQHOGNETTAGANLTWNPATVNGSISQSSTY 643
QY 797 KQ 798
DB 644 RQ 645
RESULT 15
PIP_LACLC
ID PIP_LACLC STANDARD; PRT; 1902 AA.
AC P16271;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PI-type proteinase precursor (EC 3.4.21.-) (Wall-associated serine
proteinase).
GN PRTP.
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC plasmid pMW05.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WG2;
```

RX MEDLINE=88149035; PubMed=3278687;
RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;
RT "Nucleotide sequence of the cell wall proteinase gene of
RL Streptococcus cremoris Wg2.";
CC Appl. Environ. Microbiol. 54:231-238(1988).
CC -!- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
CC GROWTH OF THE BACTERIA ON MILK.
CC -!- CATALYTIC ACTIVITY: ENDOPEPTIDASE ACTIVITY WITH VERY BROAD
CC SPECIFICITY, ALTHOUGH SOME SUBSITE PREFERENCE HAVE BEEN NOTED,
CC E.G. LARGE HYDROPHOBIC RESIDUES IN THE P1 AND P4 POSITIONS, AND
CC PRO IN THE P2 POSITION. BEST KNOWN FOR ITS ACTION ON CASEINS,
CC ALTHOUGH IT HAS BEEN SHOWN TO HYDROLYSE HEMOGLOBIN AND OXIDIZED
CC INSULIN B-CHAIN.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Cell wall.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE
CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M24767; AAA17677.1; -
DR HSP; P00782; I501.
DR MEROPS; S08.019; -
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 3.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal; Plasmid;
KW Transmembrane.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1902
FT DOMAIN 188 1876 PI-TYPE PROTEINASE.
FT TRANSMEM 1877 1895 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1896 1902 MEMBRANE ANCHOR (POTENTIAL).
FT ACT_SITE 217 217 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 281 281 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 620 620 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DOMAIN 1867 1872 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT SEQUENCE 1902 AA; 199910 MW; 2901C7F19B2E5D0B CRC64;

Query Match 3.5%; Score 146; DB 1; Length 1902;
Best Local Similarity 19.8%; Pred. No. 1.1;
Matches 176; Conservative 116; Mismatches 354; Indels 244; Gaps 45;
QY 4 SYFKGQVSANTMAYMMV-MSTHAQAADFMANDTIT-----GLQRTVIESLQSVLPRLG 58
DB 646 AYYKQLRGALTDFLTKVEMNTAQPINDYNNVIVSPRRQAGLVVDVKAIDAILEKNPS 705
QY 59 QVSENL-ADGVKALVATG-----NFSDVQVYHQEGRIIYQVTERPLIAEFEGNRLI 112
DB 706 TVVAENGYPVELKDFSTDKTLFTNSTTH-----ELTYQM-----DSNTDTNAVY 754
QY 113 -----PREGL--QEGKNAGLAVGOPLKQATVQMIELT-----NQVISQGYNTE 157
DB 755 TSATDPNSGVLYDKIDGAATKAGSNITVPAGKTAQIEFTLSLPKSPDQQQVVEGFLNFK 814
QY 158 ITVKQTMLDGNRVKLDMTFAEGKPARVVVDINIIGNQHFSDADLIDVLAIDKNKINPLSKA 217

DB 815 GS-----DGRNLNLPY-----MGFFGD--WMDGKIVDSL-----NGITYSPAG 850
QY 218 DRYTOEKLVTSLENLRAKYLNAGFV-----RFEIKDAKLNTNEDKNRIF----- 261
DB 851 GNFGTVPLLTN-KWTGTOY--GGWVTDADGQVTDQDAIAFSSDKNALINDISMKYLLR 908
QY 262 ----VEISLHEGEQVRFQGTQFLGNL--TYTQAELEALLKFAEGRFSQAMLEQTTNIS 315
DB 909 NISNVQVILDGQGNKVTTLSSTNLTKTYYNAHSQOYIYNA-PAWDTGYVDQDGNIK 967
QY 316 TKFGDDGYYIAQIRPVTRINDESRIVDEYVIDPVHPVYVRINF---TGPKTQDEVLR 372
DB 968 T--ADDSYTYRISGVPPEGDKQVDFPKLDSKAPT--VRHVALSAKTENGKTYYLTA 1024
QY 373 REMQLEGALASNQIKLSRARLMRTGFFKHVTVTRPVNPSDQVDVNFVVEEQPSGS 432
DB 1025 EAKDLSG-----LDATKSVKTAI-----NEVTNLDATF-----TDAG 1057
QY 433 TIAAGYSQSGGVTFQFVDSQNNFMGTGKHV-----NASFSRSTREYISLGMTNPFY 485
DB 1058 TTADGYTK---IETPLSDEQAQALGNGDNSAELYLTDNASNATDQDASVQKPGSTFDLI 1114
QY 486 VNG-----VSQSLSGYYRKT-----YDN-----KNISNYVLDSVGG 517
DB 1115 VNGGIPDKISSTTGYEANTQGGGYTFTSGTYPAADVDTYNTAOGKKHDLNTTYDAATN 1174
QY 518 SLSYGYPI-DENQRISFGINADNTKLHGGRFMGINSVKQLMADGGKIQVDNNGIPDFKH- 575
DB 1175 SFTASMPVYNADYAAQVDLYAD--KAHTQLLKHFDTKVLRTAPTFTDLKFNNGSDOTSEA 1232
QY 576 -----DYTTYNAILGNWYSSLDPRVPFTQGMSH-SVDLTVGFGDK----- 614
DB 1233 TIKVTGTVSADTKTVN--VGDVVAALD-----AQHHFSDVPVNYGDNITKVTATDED 1283
QY 615 ----THQKVYQGNIRP-FIKKSVLRGYAKLGYGNLPPFYENFAGGYGVRGVDQSSL 669
DB 1284 GNTTEQKTISS--YDPMKLSV-----FDQGVTFGANEF--NATSAKFYDPKT- 1331
QY 670 GPRQAYLTARRGQQTTLGEVVGNA-----LATFGSELILPLPKGDWIDQ 716
DB 1332 --GIATITGVKHPHTTTLQVDGKQIPIKDDLTFFSFTLDTLGLGK-----PF----- 1376
QY 717 VRPVIFEGGVFDTTGMDKQTDILTQFKDQATQAEQNAKANRPLLTQD 766
DB 1377 -----GVVVGDTQNKTFQEAULTFILDVAVAPTLSESSTADPVYND 1418

Search completed: September 5, 2002, 10:03:36
Job time: 529 sec

```
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE SURFACE ANTIGEN.
GN VC2252.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ercolano M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA Mcdonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL: AF004297; AAF95396.1; -.
DR TIGR: VC2252; .
DR InterPro: IPR000184; Bac_surfAg_P15.
DR Pfam: PF01103; Bac_surface_Ag; 1.
KW Complete proteome.
SQ SEQUENCE 803 AA; 89930 MW; 1998B3838F6C041D CRC64;

Query Match 28.6%; Score 1203; DB 16; Length 803;
Best Local Similarity 33.4%; Pred. No. 4.5e-58;
Matches 278; Conservative 155; Mismatches 328; Indels 72; Gaps 15;

QY 16 MAVVMVMTTHAQAD-FVANDITITGLQVVTIESLQSVLPFRLGQVSVENQADGVKALY 74
DB 8 LATLLASVANGAEKFFVVDIQIDGLQRLVALGAALLKMPVRGSDVSDQVANIYKALY 67
QY 75 ATGNFSDVQVHQEIRIIVQTERPLIAEINPEGNRLIPKEGLQGLKNAGLAVGQPLKQ 134
DB 68 SSGNFEDVKVLRDGNLTAVQVKERTIASVSFSGNKAIEEQLKQNLKLEASSIRVGEALDR 127
QY 135 ATVQMIETELNQLYISQGYNTIEIVKQTMDCNRVLDKMTFAEGKPARVVDINIIGNOH 194
DB 128 TTLSNIEKLEDFYISVGKYNATKAVVTPLPRNRADLKFFVTEGSAKIQIINFIGNOV 187
QY 195 FSDADLDVLAIDKNKI--NPLSKADRYTOEKLVTSLNLRKAYLKNAGVRFEIKDAKLN 252
DB 188 FSDELLSRFNLDVAVNWFNLFAD-DKYKKQVLADGIEALRYIYLDKGLKQVDSQVQA 246
QY 253 INEDKNRIFVEISLHEGEQYRFGQTQFLGNLTYYTQAELEALKFKAEGFSQAMLEQTIN 312
DB 247 ISPDKKGVYITLNLNEGEPTYVSKVQFRGELMGKEAFTSLIPFEITGYNGSAVTRLEE 306
QY 313 NISTFKGDDGYIAIRPVTRINDSRVDVYIYIDPHVYVVRINFTGNFKTQDEVLR 372
DB 307 SVKKVLGSESGAYPQVRIPEFDDEKQVSLVHVHVEAGKRVYVRDIREVGNNSRDEVLR 366
QY 373 REMRLEGALASNOKIQLSRLARMTGFFKHTVTRPVPNMPDQVDVNFVVEEQPSGSS 432
DB 367 REMRQWEGSWLNKDIETGKTRLNLRGFFETVEVQTVRVPGEQDQDLVYSVKEANSNV 426
QY 433 TIAAGYSGGVTFQFDVSNQNMFGTGHVNAASFRRSETREYVSLGTMNPYTFVNGVSQS 492
DB 427 NFGVGYGTEGSYFQVGLQDQNFLLSGNRVGNVAMINDYKQNLLEYRDPYNNLDGVSIG 486
QY 493 LSGYRKYKYNKNTSNVLDVSGSLSYGYPIDENQRLSFGNLNADNTKLHGFRWGI SN 552
DB 487 GKVFYNQFEASEAGIVDYTNESYGTSLWAGFPFDELNRFEPFGIGYTHNKI--GNLTPIYQ 544
QY 553 VKQLMA-----DGGKIQVDNNGIPDFKHDYTTYNAILGNWYSSLDPRVPTQGMSHS- 604
DB 545 VENFLAAQASNIDSGNLLTDD-----FDINLSWTRNRLNNSYFPTAG-NHQ R 591
```

```
QY 605 --VDLTVGFCDKTHQKVYQGNIRPFIKKS-----VLRGYAKLGYGNN-----LP 648
DB 592 AFYKMTVPGSDAQYFKLQYDVRQYFPLTKKHEFTLLRG--RLGYNGYQOTDQKONLFP 649
QY 649 FYENFYAGYGSVGRYDQSSLSLSPRS--QAYLTARRGOQTTLGEVVGGNALATFSELILP 706
DB 650 FYENFYAGGFTSLRGFGSNSAGPKAVYRDYSGSNNGSDTATDDSDVSGNATALASVELIVP 709
QY 707 LPFKGDWI--DOVRPVLIFIEGGQVFDITGMDKQID-----LTFQKDPQATAEQNAKANR 760
DB 710 TPFASEARNQIRTSIFYDMASVMDTEFDYRGKADYGNQYVYDSDP----- 756
QY 761 PLLTQDKQLRYISAGVGATWYTPIGPLSISYAKPLNKKQNDQTDVTFQIGSVF 813
DB 757 -----TNYRSSVGVALQVSPMGPLVSLAKPIKKYEGDDEEFFFTIGRTF 803

RESULT 4
Q9R2E3 PRELIMINARY; PRT; 785 AA.
AC Q9R2E3;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE HYPOTHETICAL 87.9 KDA PROTEIN.
GN YZZN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Takenoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: Analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=82059454; PubMed=6272196;
RA An G., Beadiak L., Mamelak A., Friesen J.;
RT "Organization and nucleotide sequence of a new ribosomal operon in
RT Escherichia coli containing the genes for ribosomal protein S2 and
RT elongation factor Ts.";
RL Nucleic Acids Res. 9:4163-4172(1981).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93077430; PubMed=1447125;
RA Yamanaka K., Ogura T., Niki H., Hiraga S.;
RT "Identification and characterization of the smbA gene, a suppressor of
RT the mukB null mutant of Escherichia coli.";
RL J. Bacteriol. 174:7517-7526(1992).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Smallshaw J., Kelln R.;
RT "Cloning, nucleotide sequence and expression of the Escherichia coli
RT K-12 pyrH gene encoding UMP kinase.";
RL Genetics 11:59-63(1992).
RN [6]
```

RX SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94240115; PubMed=8183897;
RA Janosi L., Shimizu I., Kaji A.;
RT "Ribosome recycling factor (ribosome releasing factor) is essential
for bacterial growth.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4249-4253(1994).
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91317739; PubMed=1860827;
RA Shimizu I., Kaji A.;
RT "Identification of the promoter region of the ribosome-releasing
factor cistron (fir)." ;
RL J. Bacteriol. 173:5181-5187(1991).
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90062117; PubMed=2684966;
RA Ichikawa S., Kaji A.;
RT "Molecular cloning and expression of the ribosome releasing factor." ;
RL J. Biol. Chem. 264:20054-20059(1989).
RN [9]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=86008268; PubMed=2995358;
RA Icho T., Sparrow C.P., Raetz C.R.H.;
RT "Molecular cloning and sequencing of the gene for CDP-diglyceride
synthetase of Escherichia coli." ;
RL J. Biol. Chem. 260:12078-12083(1985).
RN [10]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91123198; PubMed=1991717;
RA Hirvas L., Koski P., Vaara M.;
RT "The ompH gene of yersinia enterocolitica: cloning, sequencing,
expression, and comparison with known enterobacterial ompH
sequences." ;
RL J. Bacteriol. 173:1223-1229(1991).
RN [11]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91100302; PubMed=1987124;
RA Dicker I., Seetharam S.;
RT "Cloning and nucleotide sequence of the firA gene and the firA200(ts)
allele from Escherichia coli." ;
RL J. Bacteriol. 173:334-344(1991).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88058790; PubMed=2824445;
RA Crowell D., Reznikoff W., Raetz C.;
RT "Nucleotide sequence of the Escherichia coli gene for lipid A
disaccharide synthase." ;
RL J. Bacteriol. 169:5727-5734(1987).
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=88139188; PubMed=3277952;
RA Coleman J., Raetz C.;
RT "First committed step of lipid A biosynthesis in Escherichia coli:
sequence of the lpxA gene." ;
RL J. Bacteriol. 170:1268-1274(1988).
RN [15]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

RX MEDLINE=92380982; PubMed=1355089;
RA Li S., Cronan J.;
RT "The genes encoding the two carboxyltransferase subunits of
Escherichia coli acetyl-CoA carboxylase." ;
RL J. Biol. Chem. 267:16841-16847(1992).
RN [16]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90094229; PubMed=1688424;
RA Zhou Z., Syvanen M.;
RT "Identification and sequence of the drpA gene from Escherichia coli." ;
RL J. Bacteriol. 172:281-286(1990).
RN [17]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90370122; PubMed=2203971;
RA Eriani G., Delarue M., Poch O., Gangloff J., Moras D.;
RT "Partition of tRNA synthetases into two classes based on mutually
exclusive sets of sequence motifs." ;
RL Nature 347:203-206(1990).
RN [18]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Miyamoto K., Inokuchi H.;
RT "Nucleotide sequence of 5'flanking region of the ribosomal RNA gene
(rrnH) in E. coli." ;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [19]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93094132; PubMed=1459951;
RA Gervais F.G., Drapeau G.;
RT "Identification, cloning, and characterization of rcsF, a new
regulator gene for exopolysaccharide synthesis that suppresses the
division mutation ftsZ84in Escherichia coli K-12." ;
RL J. Bacteriol. 174:8016-8022(1992).
RN [20]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94124004; PubMed=7904973;
RA Allikmets R., Gerrard B., Court D., Dean M.;
RT "Cloning and organization of the abc and mdl genes of Escherichia
coli: relationship to eukaryotic multidrug resistance." ;
RL Gene 136:231-236(1993).
RN [21]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93011013; PubMed=1396599;
RA Condon C., Phillips J., Fu Z., Squires C., Squires C.;
RT "Comparison of the expression of the seven ribosomal RNA operons in
Escherichia coli." ;
RL EMBO J. 11:4175-4185(1992).
RN [22]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RA Nakayashiki T., Adachi T., Miki T., Inokuchi H.;
RT "Nucleotide sequence of the rrnH gene from Escherichia coli." ;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
RN [23]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=79012454; PubMed=358189;
RA Young R.A., Steitz J.A.;
RT "Complementary sequences 1700 nucleotides apart form a ribonuclease
III cleavage site in Escherichia coli ribosomal precursor RNA." ;
RL Proc. Natl. Acad. Sci. U.S.A. 75:3593-3597(1978).
RN [24]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93116053; PubMed=1474579;
RA Cormack R., Mackie G.;
RT "Structural requirements for the processing of Escherichia coli 5 S
ribosomal RNA by RNase E in vitro." ;

RL J. Mol. Biol. 228:1078-1090(1992).

Query Match 28.5%; Score 1197; DB 2; Length 785;
Best Local Similarity 34.4%; Pred. No. 9.3e-58;
Matches 281; Conservative 145; Mismatches 319; Indels 72; Gaps 14;

QY 35 DITITGLQRVTIESLQSIVLPRLGQVSVENQLADGVKALYATGNDSVDVVYHQEGRILYQ 94
||| ||| : : : | : : : : : ||| ||| : : : |
Db 3 DIHEFGLQAVGAALLSPMPRTGTVDNEDISNTIRALTATGFNFEDVRDLRGDTLLVQ 62
||| ||| : : : | : : : : : ||| ||| : : : |
QY 95 VTERPLAEINFEQNLRIKPEGLOBLKNAGVLAVGLOPKAQVQMTELTNOYSOGYY 154
||| ||| : : : | : : : : : ||| ||| : : : |
Db 63 VKERTIASITFSGNKSVDKMLKNLEASGVVGESELDRTIADIEKLEDFYISVGKY 122
||| ||| : : : | : : : : : ||| ||| : : : |
QY 155 NTEITVKOTMIDGNRRVKLDMTFAEGKPAPRVNDINIIGNQHFSADLIDLVAIKD-----N 209
: : : ||| : : : | : : : : : ||| ||| : : : |
Db 123 SASVAKVTVPLPRNRVLDKLKFQEGVSABEQIQINIVGNHAFETDELISHFQLRDEVPWN 182
||| ||| : : : | : : : : : ||| ||| : : : |
QY 210 KINPLSKADR-YTOEKLVTSLENRAKYNAGVFPEIKDAKLINEDKNRIFVEISLHE 268
: ||| : : ||| : : : | : : : : : ||| ||| : : : |
Db 183 VW-----GDRYQOKLAGDLETLSYLDRGYARFNIDSTQVSLTPDKKGIVTVTNITE 237
||| ||| : : : | : : : : : ||| ||| : : : |
QY 269 GEQYRFQGTFGLNITYTQAELEALKFKAEFGSOAMLEOTTNNISTKFGDDGGYYIAOI 328
||| ||| : : : | : : : : : ||| ||| : : : |
Db 238 GDQKLSVEFSGNLASHSAEEQTKEPGBELINGTKYTKWEDDIKLLGRYGAYRPV 297
||| ||| : : : | : : : : : ||| ||| : : : |
QY 329 RPVTRINDESITDVEXYIDPVHPYVVRINFITGNFKTQDEVLRREMROLEGALASNQKI 388
||| ||| : : : | : : : : : ||| ||| : : : |
Db 298 QSMPEINDADTKVLNVNVDAGNRYFYRKIREGNDTSKDALVRREMROMEGAWLGSDLV 357
||| ||| : : : | : : : : : ||| ||| : : : |
QY 389 QLSRARLMRTGFFKHVTDTRPVNSPDQVNVFVEQPSGSSTIAAGYSOSGGVTFQF 448
||| ||| : : : | : : : : : ||| ||| : : : |
Db 358 DQGERLNRLGFFETFTDTRVPGSPQDVVYVKERNTGSFNGIGYTESGSVFOA 417
||| ||| : : : | : : : : : ||| ||| : : : |
QY 449 DVSONFMGTCKHVNASFSRETRYVSIGMTNPFTYVNGVSQSLGGYYRKTXYDNKNIS 508
||| ||| : : : | : : : : : ||| ||| : : : |
Db 418 GVQODNMUGTCYAVGINTGKNDQYPAELSNTPNPTFDVGSGLGRFLYNDFOADDJUS 477
||| ||| : : : | : : : : : ||| ||| : : : |
QY 509 NYVLDSYGSLSYGPYPIDENORISFGLNADNTKLHGGRFMGISNVKQLMA-----DGKG 562
||| ||| : : : | : : : : : ||| ||| : : : |
Db 478 DYTNSYCTDVTGLFPPINEYNSLRAGLVVHNSL-----SNMQPQVAMWRILYSMGE 529
||| ||| : : : | : : : : : ||| ||| : : : |
QY 563 IQVDNNGIPDFKHDTYTTNAILGNWYSSLDPRVFTTQ--MHSVDLTGVFGDKTHQKV 620
||| ||| : : : | : : : : : ||| ||| : : : |
Db 530 HPSTSDDNSFKTDFTFN--YGVTYNKLDRGYFPTDGSRVNLTGKVTIPGSDNEYYKVT 587
||| ||| : : : | : : : : : ||| ||| : : : |
QY 621 YQGNLYRPF--IKKSLVRGAKLGYGNL----PYENFYAGYGSVREGDYQSSIGPRS 673
||| ||| : : : | : : : : : ||| ||| : : : |
Db 588 LDTATYPIDDDHKWVLGRTRWGVDGLGKGEMPFYENFYAGGSTVRFOSNTIGPKA 647
||| ||| : : : | : : : : : ||| ||| : : : |
QY 674 -----YLTARRGQQOFTL---GRVWGGNLATFGSELILPLPFKGD-WIDQ 716
||| ||| : : : | : : : : : ||| ||| : : : |
Db 648 VFEPHASNYDPDYDECATOGAKDLCKSDDAVGGNAMAVASLEFITPTFISDKYANS 707
||| ||| : : : | : : : : : ||| ||| : : : |
QY 717 VRPVIFEGGVDFDTGMKDQOTIDUTQPKPOATAEQNAKANRPLLLODKOLRYASGVG 776
||| ||| : : : | : : : : : ||| ||| : : : |
Db 708 VRTSFFWMGTVMDTNWDSSQSYGYPDYSDP-----GNIRMSAGIA 748
||| ||| : : : | : : : : : ||| ||| : : : |
QY 777 ATWTPTGPLSISYAKPLNKKNDQTDIVQFOIGSVF 813
||| ||| : : : | : : : : : ||| ||| : : : |
Db 749 LOWMSPLGLPFSYAQPQKYDGDKAEQFQFNIGTW 785
||| ||| : : : | : : : : : ||| ||| : : : |

RESULT 5
Q9PEI2 ID PRELIMINARY; PRT; 784 AA.
AC Q9PEI2
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OUTER MEMBRANE ANTIGEN.
GI XF1046.
OS Xylella fastidiosa.

RT "Complete genomic sequence of Pasteurella multocida Pm70. ";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006235; AAK04076.1; -.
DR InterPro; IPR000184; Bac_surfaAg_D15.
DR Pfam; PF01103; Bac_surface_Ag_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 791 AA; 87761 MW; 024550DD8F99CD82 CRC64;

Query Match 25.0%; Score 1049; DB 16; Length 791;
Best Local Similarity 31.3%; Pred. No. 1.3e-49;
Matches 258; Conservative 162; Mismatches 344; Indels 60; Gaps 17;

QY 14 MTMAYMMVNSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSVENQADGVKAL 73
DB 4 LLIASLLFGSTTAFAPFVVKDIRVGVQAGTEGSLVATLPVRGQRTDNDIANVVRKL 63
QY 74 YATGNFSDVQVYHQGRRIYQVTERPLIAEFNFGNRLIPKEGLQGLKNAGLAVGQPLK 133
DB 64 FLSCQYDDVKASREGNTLVVTPMPKPVISNVVIDGNKSIDPAIKQNLKDANGFKVGDVLN 123
QY 134 QATVOMIETELTNQISQGYNTETTYKQTMLDGNRVKLDMTFAEGKPARVVDINIIGNQ 193
DB 124 RAKLEEFKRGIVEHNSVGRYNAKVEALVNTLPNNSAEIKIQINEDDDVALFKEITFEGN 183
QY 194 HFSADLDIVLAIKDNKINPLSKADRYTOEKLVTSLNLRKYLINAGVFRFEIKDAKLN 253
DB 184 AFSSKLEQDQELQTDAMWKL-F-GNKFQDTQFNKDLTFLRSYLDLRYGVAQFIIDTDIKL 242
QY 254 NEDKNRIFVELSLHEGEQYRFGQTFGLNLTQAELEALLKFAERG-FSQAMLEOTTN 312
DB 243 SDDKEARVILKVGEDLYTVKSARILGVDGGMGAELAPILDAIQLNGLFRANVLEVEQ 302
QY 313 NISTKFGDGGYYIAQIRPVRTRINDESRVDVEYIDPVHPVYVRRINFTGNFKTQDEVLR 372
DB 303 RIKSKLGERGYATAQVNVHPFDDQDKTISLDFIVEAGKSYTVQIRFEGNTSSADSTLR 362
QY 373 REMRQLEGALASNOKIQLSRARLMRTGFHKHVTVDTRPVNSPDQVDVNFVEPQSGSS 432
DB 363 QEMRQEQEAGWLSSELVELGLRLDRTGFFESVETKTEAIPGS-DQVDVIYKVERNTGSI 421
QY 433 TIAAGYSQSGGVTFQFDVSONNEMGTCKHVNASFSSRETREYVSLGNTNPFYTVNGYSQS 492
DB 422 NFGIGYTESGLSYQASIKQDNFLGMGSSISLGGTRNDYGTINLGYNEPYFTKDGV--S 479
QY 493 LSGYYRKTKYDNKNISN---YVLDSYGSLSYGYPIDENQRISEGLNADNTKLHGGRFMG 549
DB 480 LGGNVFEEYDSSKSNSTSAAYGRTSYGNLTGLFPVNNENSYILGVGYTNKL----- 532
QY 550 ISNV-KOLMADGGKIQVNDNGIPDFK-HDYTTYNAILGWNYSSLDRVPFPTQGMSSHVD- 606
DB 533 -KNIAPEYNDLYRQSMKYNDWSWTFKSHD---FDLSFGWYNSLNRGYFPTKGRANIGG 588
QY 607 -LTVGFGDKTHQVYVYVGNVYRPF-----IKKSVLRKYLKGYGNLPPFENFY 654
DB 589 RVLIIPSDNKKYKLNAAQAGFYPLDRHGWVLSRSISASFADFG-----GKRLPFYQYIS 644
QY 655 AGGYSGVRGYDQSSGLSPRSQAYLTARRQOQTLT--GEVGGNALATFGSELILPLPFKGD 712
DB 645 AGGIGSLRGFAYGAIGPNA-IYRTRQCPDYSCLVSSDVGIGNAMVASTELLIVPTFVAD 703
QY 713 W-IDVQRPVIFIEGGQVDFDTGMDKQTDITLQFKDQATAEQNAKAA--NRPLLTQDKQL 769
DB 704 KNQNSVTSLSFVDAASWNT-----RWKAEDKAKFAKLVNPDYSDPSRV 747
QY 770 RYSAGVATWYTPIGPLSISYAKPLNKKONDQTDYTVQFQIGSVF 813
DB 748 RASAGVALQMSPIGPLVFSYAKPLKYYQGEDEIEQFQFSGTTF 791

RESULT 13

Q51930
ID Q51930 PRELIMINARY; PRT; 789 AA.
AC Q51930; Q51921;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE OUTER MEMBRANE ANTIGEN OMA87.
GN OMA87.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PBA100;
RA Ruffolo G.C.; Adler B.;
RL Infect. Immun. 64:0-0(0).
RN [2]
RP SEQUENCE OF 750-789 FROM N.A.
RC STRAIN=9222;
RX MEDLINE=95369730; PubMed=7642134;
RA Delanarche C.; Manoha F.; Behar G.; Houlgatte R.; Hellman U.;
RA Wroblewski H.;
RT "Characterization of the Pasteurella multocida skp and fira genes."
RL Gene 161:39-43(1995).
DR EMBL; U60439; CAA44600.1; -.
DR EMBL; X74357; AAC52399.1; -.
DR InterPro; IPR000184; Bac_surfaAg_D15.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF01103; Bac_surface_Ag_1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 789 AA; 87437 MW; DC1D004AAF2CD9E CRC64;

Query Match 23.8%; Score 1000; DB 2; Length 789;
Best Local Similarity 30.8%; Pred. No. 6.4e-47;
Matches 254; Conservative 164; Mismatches 343; Indels 64; Gaps 21;

QY 14 MTMAYMMVNSTHAQAADFMANDITITGLQRTVIESLQSVLPFRLGQVSVENQADGVKAL 73
DB 4 LLIASLLFGSTTAFAPFVVKDIRVGVQAGTEGSLVATLPVRGQRTDNDIANVVRKL 63
QY 74 YATGNFSDVQVYHQGRRIYQVTERPLIAEFNFGNRLIPKEGLQGLKNAGLAVGQPLK 133
DB 64 FLSCQYDDVKASREGNTLVVTPMPKPVISNVVIDGNKSIDPAIKQNLKDANGFKVGDVLN 123
QY 134 QATVOMIETELTNQISQGYNTETTYKQTMLDGNRVKLDMTFAEGKPARVVDINIIGNQ 193
DB 124 RAKLEEFKRGIVEHNSVGRYNAKVEALVNTLPNNSAEIKIQINEDDDVALFKEITFEGN 183
QY 194 HFSADLDIVLAIKDNKINPLSKADRYTOEKLVTSLNLRKYLINAGVFRFEIKDAKLN 253
DB 184 AFSSKLEQDQELQTDAMWKL-F-GNKFQDTQFNKDLTFLRSYLDLRYGVAQFIIDTDIKL 242
QY 254 NEDKNRIFVELSLHEGEQYRFGQTFGLNLTQAELEALLKFAERG-FSQAMLEOTT 311
DB 243 SDDKKEPCL-IS-EGDLYTV-KTRVSGGMWGMGAELAPILETIQNLGLFRSTSVLEVE 299
QY 312 NNISTKFGDGGYYIAQIRPVRTRINDESRVDVEYIDPVHPVYVRRINFTGNFKTQDEV 371
DB 300 QNKKSLGERGYATAQVNVHPHTDEQDKTISLDFIVEAGKSYTVRQIRFEGNTSSADSTL 359
QY 372 REMRQLEGALASNOKIQLSRARLMRTGFHKHVTVDTRPVNSPDQVDVNFVEEOPSGS 431
DB 360 QEMRQEQEAGWLSSELVELGLRLDRTGFFESVETKTEAIPGS-DQVDVIYKVERNTG 418
QY 432 SITTAAGYSQSGGVTFQFDVSONNEMGTCKHVNASFSSRETREYVSLGNTNPFYTVNGV 491
DB 419 INFGIGYTESGLSYQASIKQDNFLGMGSSISLGGTRNDYGTVNLGYNEPYFTKDGV-- 476
QY 492 SLSGYYRKTKYD---NKNISNVYLDYSGSSLSYGYPIDENQRISEGLNADNTKLHGGRFM 548

```
Db 477 SLGGNVSEEDSSKSNSTASGYGRTSYGNNLTGLPPVNNNSYILGVGYTNKL----- 530
Qy 549 GISNV-KOLMADGGKQVNDNGIPDK-HDYTTYNAILGWYSSLDPRVFTQGHSHVD 606
Db 531 --LTAPEYNDLYRQSMKYNDSWTFKSHD--FDLSFGWYNSLNRGYFPTKGVANIG 585
Qy 607 --LTVFGDKTHQKVYVYOGNYRPP-----IKSVLRGVAKILGYGNLPPFYENF 653
Db 586 GRVTIPGSDNYYKLNABOAGFYPLDRHGWLSRSISASADGFS---GKRLPFYQYY 641
Qy 654 YAGGVYRGVYDQSSGLSPRSQAYLTARGQQTTL--GEVVGSNALATFGSELILPLPKG 711
Db 642 SAGGIGSURGAYGAIGNA-IYRQCPDSCYCLVSSVDVIGNAMVTADELIVPTPPVA 700
Qy 712 DW-IDQVRPVIIEGGQVFTDTGMDKQIDLTQFKDPQATAEQNAKAA--NRPLLTQDKQ 768
Db 701 DKNQSVRTSLFVDAASVWNT-----RWKAEDRAKFAKLVNPDYDSFR 744
Qy 769 LRYSGAGCATWTPPIGPLISYAKPLNKKQNDQDTTVQFGSVF 813
Db 745 VRASGVALQWQSPITGLVFSYAKPLKRYQGEIQEIQFQFSGTGF 789

RESULT 14
Q9ZE03
ID Q9ZE03 PRELIMINARY; PRT; 768 AA.
AC Q9ZE03;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE OUTER MEMBRANE PROTEIN OMP1 (OMP1).
GN RPI60.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=792;
RN [1]
RC STRAIN=MADRID E;
RA Anderson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
DR EMBL: AJ235270; CAAL4627.1;
DR InterPro: IPR00184; Bac_surfaG_d15.
DR Pfam: PF01103; Bac_surfaceAg; 1.
KW Complete proteome.
SQ SEQUENCE 768 AA; 87196 MW; BIAB12D783D9FDCB CRC64;

Query Match 17.98; Score 754; DB 16; Length 768;
Best Local Similarity 26.38; Pred. No. 2.1e-33;
Matches 216; Conservative 145; Mismatches 383; Indels 76; Gaps 20;

Qy 7 KGFOVSAMTAMVMVMSTHQAQAFMANDITITGLQRTVIESLOSVLPERLQGVYSENL 66
Db 2 KIISIKTLILLIFYYHIFSAFVIVIKITIEGNHRVERSTIESYLNKYGEYNNKS 61
Qy 67 ADGVKALYATNFSDVQVY-HQEGRIIVQVTERPLIAINEGRNLIPKEGLQEGKNAG 125
Db 62 DEAIKRLYATSLFRININMYITDGNLIVNVTETPISSVWFSGSKIKTNILAKEIYMS 121
Qy 126 LAVGQPLQAQVOMIETELTNQYISQGYNTIETVQTMLDGNRVKLDMTFAEKGPARV 185
Db 122 ---GESLSQAQIELDVKKILEIYRSRGFSTKVPKIKSLNNRVKVIIDFAEGPKTVIK 178
Qy 186 DINIIGNOHFSDADIDVLATKDNK-INPLSKADRYTOEKLVTSLENIRAKYLVNAGVRF 244
Db 179 SIYFSGNEHYSDELKSLVLTKEKNWFRFLESNDTIDPRVEYDEKELLREFYQSVGFAD 238
```

```
Qy 245 EIKDAKLINEDKNRIFVEISLHEGEQVRFQGTQFLGNLTYYTQ-AELEALLKFKAEEGFS 303
Db 239 RVISASVALNDTKKEYTFTTISTEEGEKYRFGNVTIDNKLNTINIKQLNIVIKQCKIEN 298
Qy 304 QAMLEQTTNNISTKGGDDGYAAQIRPVTRINDESRITVDVEYIIDPVHPVYVRRINFTCN 363
Db 299 MKTVDDIAEKIGEXTANGYPANVYIPDKND-NHTADIKFPIIEKADKVYINKININN 357
Qy 364 FKTDQEVLRREMRQLEGALASNOKIOLSRARLMRTGFFKHVTVDTRPVNSPDQVDNV 423
Db 358 LKTEDHVIIRAFKIEGDMNRSYIEKGERNLRLNIDFEKVSISLAQT-KADKKYDVNVE 416
Qy 424 VEEQPSGSTIAAGYSQSGVTFQFDVSONPMGTGKHVNASFSESTREYVSLGMTNYP 483
Db 417 VDEKSTSTGFDLGYNTAGGLFGREFLERNLVGTGKLILNAGVQVSKNSTSYGCTIDPH 476
Qy 484 FTVNGVSQSLSGY---RKTKYDNKNISNYVLDYSYSGSLSYGYPIDE--NQRISEGLNA 537
Db 477 FLDRDLSUSVNAFRNYTGRGASVLNTDQSYKLSIGVKISLGYDMKEDLSHEIDYLIKR 536
Qy 538 DNTKLHGRFMSGISNVKQLMADGGKIQVDNNGIPDFKHDTTYNAILGNWYSSLDPRVPP 597
Db 537 DILSAPSP-----SNSIFLNEQMGKLITSAIG-----HTIT-----YQTDNKIYP 577
Qy 598 TQG--MHSVDLTVGFGDKTHQVYVYQGYNIYRPFIKKSVLRGYAKLG-----YGNLPP 649
Db 578 KNGYLVSGTQEFAGVGGDNKYIKHEIECKFYKFINNKVTKLSAAGGDMAGLGKMWRI 637
Qy 650 YENFYAGGVSGVYDQSSLPSPRSQAYLTARGQQTTLGEVVGGNALATFGSELILPLPF 709
Db 638 SDRFNLDY-SLRGFASGGVGPR-----EKNTEGILGGERYYTFTSELNFPPTPV 685
Qy 710 KGDIDQVRPVIFIEGGQVFTDGTMDKQIDLTQFKDQATAEQNAKAAANRLLTQDKOL 769
Db 686 PEEF--NFTGAVFIDLGSVWG-VGLNKK-----QYKTPNG-----FYNDQSL 724
Qy 770 RYSAGVGATWTPPIGPLISYAKPLNKKQNDQDTTVQFQI 809
Db 725 RASVGFGFIWVTRFAPIRMDWGFPIKKQYD--DTQNFHL 762

RESULT 15
Q92Q48
ID Q92Q48 PRELIMINARY; PRT; 776 AA.
AC Q92Q48;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE OUTER MEMBRANE TRANSMEMBRANE PROTEIN.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RC SEQUENCE FROM N.A.
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Goudie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelie D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebault P., Vandenbol M.,
RA Vothoelster F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR EMBL: AL591787; CAC46081.1;
KW Complete proteome.
SQ SEQUENCE 776 AA; 84520 MW; CE261E6372A94726 CRC64;
```

Search completed: September 5, 2002, 10:02:56
Job time: 530 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:13:56 ; Search time 61.4 seconds
(without alignments)
1470.733 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYKFGQVSAMTMMVM.....LNKKQNDTDTVOFGISVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 20

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	813	100.0	813	21	AYY44390
2	749	92.1	813	21	AYY44391

ALIGNMENTS

RESULT 1
AYY44390
ID AYY44390 standard; Protein: 813 AA.
XX

AC AAY44390;
XX
XX 14-MAR-2000 (first entry)
XX
XX M. catarrhalis BASB027 polypeptide #1.
XX
XX BASB027; OMP85; outer membrane protein; otitis media; treatment;
KW diagnosis; bacterial infection.
XX
XX Moraxella catarrhalis.
OS
XX WO9963093-A2.
PN
XX 09-DEC-1999.
PD
XX
XX 31-MAY-1999; 99WO-EP03822.
PF
XX
XX 03-JUN-1998; 98GB-0011945.
PR
XX 08-MAR-1999; 99GB-0005304.
PR
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA
XX Vinals-Bassols C;
PI
XX WPI; 2000-105700/09.
XX
XX N-PSDB; AA229550.
XX
XX Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media
XX
XX Claim 1; Fig 3; 109pp; English.
XX
XX The present sequence is BASB027 polypeptide, which shows significant
CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is
CC encoded by DNA obtained from chromosomal DNA library of Moraxella
CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and
CC polypeptide can be used for diagnosis and staging of disease, determining
CC susceptibility to a disease and to prepare medicaments for treating M.
CC catarrhalis infections, especially otitis media. The BASB027 DNA can be
CC used as probe for screening of genetic mutations, serotype, taxonomic
CC classification or identification. BASB027 agonists, antagonists and
CC antibodies may be used to prevent and/or treat bacterial infections.
XX
XX Sequence 813 AA;
XX
XX Query Match 100.0%; Score 813; DB 21; Length 813;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRNSYKFGQVSAMTMMVMVMSTHAQAADFMANNDITITGLQRTVTTESLQSVLPFRLQGV 60
Db 1 mrnsyfkfgqvssamtmmvmsthaqaadfmndititglrvrtieslqsvlpfqlgv 60
QY 61 VSENLADGVKALYATGNFSDVQVYHQEGRIITYQVTERPLIAEINFEGNRLIPKEGLQEG 120
Db 61 vseqladgvkalyatgnfsvdvqvyhqegriilyqvtterpliaefnegrllipkeglqeg 120
QY 121 LKNAGLAVGQPLKQATVQMTELTINQYISQGYNTETITVKTQMLDGNRVKLDMTFAEGK 180
Db 121 lknaglavgqplkqatvqmteletnqyisqgyntetivktqtmldgnrvkldmtfaegk 180
QY 181 PARVVVDINIGNHQFSDADLIDVLAIKDNKINPLSKADRYTQEKLVTSLENIRAKYLNAG 240
Db 181 parvvvdinignhfsdaddlidvlaikdnklnplsksdrytqeklvtslenirakylng 240
QY 241 FVRFEIKDAKLNINEDKNRIFVEISLHEGEQYRFQGTQFLGNLTYYTQAEALLKFAEE 300
Db 241 fvrfeikdaklninedknrfveislshegeyrfqgtqflgnltytqaeallkfaee 300
QY 301 GFSQAWLEQTTNNISTKFGDDGYYYAQIRPVTRINDESRVDEYIDVPVHVYVRINF 360
Db 301 gfsqawleqttnnistkfgddgyyyaqirpvtrindestrvdeyidvpvhvyvrinf 360

Db 301 gfsqamleqttnnistkfgdgyyaaqrprvtrindesrtvdveyyidpvhpyvrrinf 360
QY 361 TGNFKTQDEVLREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNPSDQVDV 420
Db 361 tgnfktqdevlremrqlqegalsnkiqlsrarlmrtgffkhvtvdtrpvnsdpqvdv 420
QY 421 NFVVEQPSGSSTIAAGYSQSGGVTFQFDYSQNNFMGTGKHVNASFSSRSETREYISLGMT 480
Db 421 nfveeqpsgstiaagysqsggvtfqfdvsnqnmfngtghvnasfssrsetrevyslgmt 480
QY 481 NPFTVNGVSQSLSGYRKTKYDNKNISNVYLDYSYGGSLSYGYPIDBNQRIISFGLNADNT 540
Db 481 npftvngvsqslsgyrrtktdnknisnvyldysygslygyypidenqrtsfglnadnt 540
QY 541 KLHGGRFMGTSNVKQJLMDADGKIQVDNNGIPDFKHDTTYNAIILGWYSSSLDRPVPTQG 600
Db 541 klhggrfmgtsnvkqlmadgsklqvdnngipdfkhdyttnaillgnwssldrppvptqg 600
QY 601 MSHSVDLTVGFGDKTHQVYQYQNIYRPFIFKSVLRGYAKLGYGNLPPFYENFYAGGYGS 660
Db 601 mshsvdltvfgdkthqkvyyqgniyrfpikksvlgryaklgygnlppfyenyfaggygs 660
QY 661 VRGYDSSLGPRQAYLTARRGGQTLGEGVVGGNALATFGSELILPLPFGDWIDQVRPV 720
Db 661 vrgydgsslgprsqayltarrgqgttlgevvvggnalatfgselilplpfkgdwidqvrpv 720
QY 721 IFTEGGQVFTTGMKDQITDLTQFKDQATAEQNAKAANRPLLTQDKOLRYSGAGVATWY 780
Db 721 ifteggqvfttgmkdqitdltfkdqataeqnakaanrplltqdkqlrysagvatwy 780
QY 781 TPIGPLSISYAKPLNKKQNDQDTVQFQIGSVF 813
Db 781 tpigplsisyakplnkkqndqdtvqfignsvf 813

RESULT 2
AA144391
ID AAY44391 standard; Protein; 813 AA.
XX
AC AAY44391;
XX
DT 14-MAR-2000 (first entry)
XX
DE M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.
XX
KW BASB027: OMP85; outer membrane protein; otitis media; treatment;
KW diagnosis; bacterial infection.
XX
OS Moraxella catarrhalis.
XX
PN WO9963093-A2.
XX
PD 09-DEC-1999.
XX
PF 31-MAY-1999; 99WO-EP03822.
XX
PR 03-JUN-1998; 98GB-0011945.
PR 08-MAR-1999; 99GB-0005304.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Vinals-Bassols C;
PI
DR WPI: 2000-105700/09.
DR N-PSDB; AA229551.
XX
PT Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media
XX
PS Claim 1; Page 102-104; 109pp; English.
XX
XX The present sequence is BASB027 polypeptide, which shows significant

CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is
CC encoded by DNA obtained from chromosomal DNA library of Moraxella
CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and
CC polypeptide can be used for diagnosis and staging of disease, determining
CC susceptibility to a disease and to prepare medicaments for treating M.
CC catarrhalis infections, especially otitis media. The BASB027 DNA can be
CC used as probe for screening of genetic mutations, serotype, taxonomic
CC classification or identification. BASB027 agonists, antagonists and
CC antibodies may be used to prevent and/or treat bacterial infections.
XX
SQ Sequence 813 AA;

Query Match 92.1%; Score 749; DB 21; Length 813;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 QLADGVKALYATGNFSDVQVYHQEGRIYQVTERPLIAEINFEGNRLIPKEGLQGLKNA 124

Db 65 qladgvkalyatgnfsvdqvyhqegriiyqvtterpliaeinfegnrlipkeglqglkna 124

QY 125 GLAVGQPLKQATVQMIETELTNQYISQGYNTETITVQTMLDGNRVKLDMTFAEGKPARV 184

Db 125 glavgpqlkatvqmieteltnqyisqgyntetitvqtmldgnrvkldmtfaegkparv 184

QY 185 VDINIIGNOHFSDADLDVLAIKDNKINPLSKADRYTOEKLVTSLLENRAKYLNAGVRF 244

Db 185 vdnignohfssadldvlaikdnknplskadrytqeklvtslenlrakylnagvrf 244

QY 245 EIKDAKUNINEDKNRIFVEISLHGEQYRFGQTQFLGNLYTQAELEALUKFAKEEFGSQ 304

Db 245 eikdakuninedknrfveislhgeqyrfgtqflgnlytqaeleallkfkaeefsq 304

QY 305 AMLEQTTNNISTKFGDDGYVYQAQIRPVTRINDESRTVDVEYIDVHPVYVRRNFTGNF 364

Db 305 amleqttnnistkfgdgyyaaqrprvtrindesrtvdveyyidpvhpyvrrinf 364

QY 365 KTQDEVLRREMRQLEGALASNOKIQLSRARLMRTGFFKHVTVDTRPVNPSDQVDVNFV 424

Db 365 ktqdevlrremrqlqegalsnkiqlsrarlmrtgffkhvtvdtrpvnsdpqvdvnf 424

QY 425 EEPQSGSTIAAGYSQSGGVTFQFDVSQNNFMGTGKHVNASFSSRSETREYISLGMTNYPF 484

Db 425 eeqpsgstiaagysqsggvtfqfdvsnqnmfngtghvnasfssrsetrevyslgmt 484

QY 485 TVNGVQSLSGYRKTKYDNKNISNVYLDYSYGGSLSYGYPIDENQRIISFGLNADNTLHG 544

Db 485 tvngvsglsgyrrtktdnknisnvyldysygslygyypidenqrtsfglnadntl 544

QY 545 GRFMGISNVKQLMADGKIQVDNNGIPDFKHDTTYNAIILGWYSSSLDRPVFPPTQGMSHS 604

Db 545 grfmgisnvkqlmadgsklqvdnngipdfkhdyttnaillgnwssldrppvfpqgmshs 604

QY 605 VDLTVGFGDKTHQVYQYQNIYRPFIFKSVLRGYAKLGYGNLPPFYENFYAGGYSGVRGY 664

Db 605 vdltvfgdkthqkvyyqgniyrfpikksvlgryaklgygnlppfyenyfaggy 664

QY 665 DQSSSLGPRQAYLTARRGGQTLGEGVVGGNALATFGSELILPLPFGDWIDQVRPVFIE 724

Db 665 dqsslgrsqayltarrgqgttlgevvvggnalatfgselilplpfkgdwidqvrpv 724

QY 725 GGQVFDFTTGMKDQITDLTQFKDQATAEQNAKAANRPLLTQDKOLRYSGAGVATWYPIG 784

Db 725 ggqvfdfttgmkdqitdltfkdqataeqnakaanrplltqdkqlrysagvatw 784

QY 785 PLSISYAKPLNKKQNDQDTVQFQIGSVF 813

Db 785 plsisyakplnkkqndqdtvqfignsvf 813

Search completed: September 5, 2002, 10:13:58
Job time: 386 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:14:27 ; Search time 25.27 Seconds
(without alignments)
785.833 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 2442594 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
- 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description

No matches found

Search completed: September 5, 2002, 10:14:27
Job time: 285 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:15:14 ; Search time 39.83 Seconds
(without alignments)
1961.351 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQTDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
-----	-----	-----	-----	-----	-----	-----

No matches found

Search completed: September 5, 2002, 10:15:14
Job time: 242 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:22:07 ; Search time 20.99 Seconds
(without alignments)
1499.714 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTMAVM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
---------------	-------	----------------	--------	-------	-------------

No matches found

Search completed: September 5, 2002, 10:22:07
Job time: 454 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:21:40 ; Search time 63.52 Seconds
(without alignments)
2214.183 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

- Database : SPTREMBL_19:**
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mhc:*
 - 8: sp_organelle:*
 - 9: sp_phage:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_rvirus:*
 - 16: sp_bacteriap:*
 - 17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

No matches found

Search completed: September 5, 2002, 10:21:40
Job time: 457 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:04:37 ; Search time 61.46 Seconds
(without alignments)
1469.297 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTMMVM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 15

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : A_Geneseq_032802:*

1: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*

2: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*

3: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*

4: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*

5: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*

6: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*

7: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*

8: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*

9: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*

10: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*

11: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*

12: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*

13: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*

14: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*

15: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*

16: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*

17: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*

18: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*

19: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*

20: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*

21: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*

22: /SIDSL1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total of the result being printed, and is derived by analysis of the score distribution.

SUMMARIES				ALIGNMENTS	
Result No.	Score	Query Match	Length DB ID	Description	
1	813	100.0	813 21	AA1980	M. catarrhalis BAS
2	749	92.1	813 21	AA1981	M. catarrhalis (AT
3	16	2.0	17 21	AA1983	Moraxella catarrha
4	15	1.8	16 21	AA1984	Moraxella catarrha

ID	AA1980	standard; Protein; 813 AA.
XX	AA1981	
AC	AA1982	
XX	AA1983	
DT	14-MAR-2000	(first entry)
XX		
DE	M. catarrhalis BAS027	polypeptide #1.
XX		
DE	BAS027; OMP85;	outer membrane protein; otitis media; treatment;
KW	diagnosis; bacterial	infection.
XX		
OS	Moraxella catarrhalis.	
XX		
PN	WO9963093-A2.	
XX		
PD	09-DEC-1999.	
XX		
PF	31-MAY-1999;	99WO-EP03822.
XX		
PR	03-JUN-1998;	98GB-0011945.
PR	08-MAR-1999;	99GB-0005304.
XX		
XX	(SMIK)	SMITHKLINE BEECHAM BIOLOGICALS.
PA		
XX	Vinals-Bassols C;	
PI		
XX		
DR	WPI; 2000-105700/09.	
DR	N-PSDB; AA229550.	
XX		
PT	Novel BAS027 polynucleotide and polypeptides from Moraxella	
PT	catarrhalis useful for treating M. catarrhalis infection such as otitis	
XX	media -	
PS	Claim 1; Fig 3; 109pp; English.	
XX		
CC	The present sequence is BAS027 polypeptide, which shows significant	
CC	homology to Neisseria meningitidis OMP85 outer membrane protein. It is	
CC	encoded by DNA obtained from chromosomal DNA library of Moraxella	
CC	catarrhalis strain Mc2931 (ATCC 43617). BAS027 polynucleotide and	
CC	polypeptide can be used for diagnosis and staging of disease, determining	
CC	susceptibility to a disease and to prepare medicaments for treating M.	
CC	catarrhalis infections, especially otitis media. The BAS027 DNA can be	
CC	used as probe for screening of genetic mutations, serotype, taxonomic	
CC	classification or identification. BAS027 agonists, antagonists and	
CC	antibodies may be used to prevent and/or treat bacterial infections.	
XX		
SQ	Sequence	813 AA;
Query Match 100.0%; Score 813; DB 21; Length 813;		
Best Local Similarity 100.0%; Pred. No. 0;		
Matches 813; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	MRNSYFKGFQVSAMTMMVMVSTHAQAADFANDITITGLQRTVTSQSULPRLGQV 60
Db	1	mrnsyfkqvsamtmvmvsthaqaadfmndititglqrvtiesqsvlpfqlgv 60
QY	61	VSENLADGVKALYATGNFSDQVYHQEGRIVQVTERPLIAEINFEGRNLPKEGLQEG 120
Db	61	vseqladgvkalyatgnfsdqvyhqegrivqvterpliaeinfeqrnlipkeglqeg 120
QY	121	LKNAGLAVGQPLKQATVQMIETELTNQYISQGYNTEITVTKQTMLDGNRVKLDMTFAEGK 180
Db	121	lknaglavgqplkqatvqmieteltnqyisqgynteitvktqmdgnrvkldmtfaegk 180
QY	181	PARVVDINIIQNHQSFADLIDVLAIKDNKNIPSLKADRYTQEKLVTSLENRAKYLNAG 240
Db	181	parvvdiniinqhfsdaldidvlaikdnknipslkadytqeklvtslenlrakylnaag 240
QY	241	FVRFELKDAKLNINEDKNRIFVEISLHGEQYRFGOTQFLGNLTYYTQAEALLLKFAEE 300
Db	241	fvrfeikdaklninedknrifveislhegeyrfgqtqfignlttytqaealeallkfaee 300

301 GFSQAMLEQTTNNISTKFGDDYYIAQIRPVTRINDESRTVDVEYIDPVHPVYRRINF 360
 301 gfsqamleqttnnistkfgddyyiaqirpvrtrindestrtdvveyidpvhpyrrinf 360
 361 TGNFKTQDEVLRMRQLEGALASNQIKQLSRARLMRTGFFKHHVTVTRPVNPSPDQVDV 420
 361 tgnfktqdevlrmerqlegalasnqiklsrarlmrtgffkhhvtvtrpvnpdpqvdv 420
 421 NFVVEEQPSGSTTAAGYSQSGVTFQFDVSONNFMCTGKHVNASFSESTREYVSLGMT 480
 421 nfveeqpsgsttaagysqsgvtfqfdvsonnfmctgkhvnasfsestreyvslgmt 480
 481 NPYFTVNGVSQSLGYRKTXYDNKNISNYVLDYSGGSLSYGYPIDENQRIISFGLNADNT 540
 481 npyftvngvsqslgyrktxydnknisnyvldysggslsygypidenqrifsglnadnt 540
 541 KLHGGRFMGINSVKOLMADGGKIQVDNNGIPDFKHDTYTYNAILGWNYSLSLRPVFTQG 600
 541 klhggrfmginvskolmadggkiquvdnngipdfkhdtytyynailgwnyslsldrvpftqg 600
 601 MSHSVDLTVGFGDKTHOKVYVQGNIRYRPFIKKSVLRGVAKLGYGNLPPFENYVAGGYGS 660
 601 mshsvdltvvgfkdthokvyvqgniryrfikksvlrgvakiigygnlppfenyvaggygs 660
 661 VRGYSQSLGPRSOAYLTARRGQTTLGEVYVGGNALTFFGSELILPLPFKGDWIDQVRPV 720
 661 vrgysqslgprsqayltarrgqttlgevvyggnaltffgselellplpfkgdwidqvrpv 720
 721 IFIEGGQVFTTGMKDQIFDLTQFKDPOATAEQNAKANRPLLTQDKOLRYSGAGVATWY 780
 721 ifieggqvfttgmkdqifdltdqfkdpqataeqnakanrplltqdkqlrysagvgatwy 780
 781 TPIGPLSISYAKPLNKKONDQTDVTFQIGSVF 813
 781 tpiGPLSISYAKPLNKKONDQTDVTFQIGSVF 813

RESULT 2

AA44391
 ID AA44391 standard; Protein; 813 AA.
 XX
 AC AA44391;
 DT 14-MAR-2000 (first entry)
 DE
 DE M. catarrhalis (ATCC 43617) BASB027 polypeptide #2.
 KW BASB027; OMP85; outer membrane protein; otitis media; treatment;
 KW diagnosis; bacterial infection.
 OS Moraxella catarrhalis.
 PN W09963093-A2.
 XX
 PD 09-DEC-1999.
 XX
 PF 31-MAY-1999; 99WO-EP03822.
 XX
 PR 03-JUN-1998; 98GB-0011945.
 PR 08-MAR-1999; 99GB-0005304.
 XX
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PA
 XX Vinals-Bassols C;
 PI WPI; 2000-105700/09.
 DR N-PSDB; AA229551.
 XX
 PT Novel BASB027 polynucleotide and polypeptides from Moraxella
 PT catarrhalis useful for treating M. catarrhalis infection such as otitis
 PT media
 XX
 PS Claim 1; Page 102-104; 109pp; English.

XX The present sequence is BASB027 polypeptide, which shows significant
 CC homology to Neisseria meningitidis OMP85 outer membrane protein. It is
 CC encoded by DNA obtained from chromosomal DNA library of Moraxella
 CC catarrhalis strain Mc2931 (ATCC 43617). BASB027 polynucleotide and
 CC polypeptide can be used for diagnosis and staging of disease, determining
 CC susceptibility to a disease and to prepare medicaments for treating M.
 CC catarrhalis infections, especially otitis media. The BASB027 DNA can be
 CC used as probe for screening of genetic mutations, serotype, taxonomic
 CC classification or identification. BASB027 agonists, antagonists and
 CC antibodies may be used to prevent and/or treat bacterial infections.
 XX
 SQ Sequence 813 AA;

Query Match 92.1%; Score 749; DB 21; Length 813;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 65 QLADGVKALYATGNFSDVYHQEGRIYQVTERPLIAEINFEGRNLRIPKEGLQELKNA 124
 Db 65 qladgvkalyatgnfsdvvyhqegriyqvterpliaeinfegnrlipkeglqelkna 124
 Qy 125 GLAVGQPLKQATVQMIEFELTNQYISQGYNTEITVKOTMLDGNVTKLDMTEAGKPARV 184
 Db 125 glavgqplkqatvqmeteltnqyisqgynteitvkqtmldgnrvtkldmtfaegkparv 184
 Qy 185 VDINIIGNOHTSDADLIDVLAIKDNKINPLSKADRYTOEKLVTSLLENRAKYLNAGFVR 244
 Db 185 vdiniignohfsdadlidvlaikdnkinplskadrytqeklvtslenlrakylnagfvrf 244
 Qy 245 EIKDAKLINEDKNRIFVEISLHEGEQYRFGQTQFLGNLTYYTQAELEALLKFAEGFSQ 304
 Db 245 eikdaklinedknrifveislhegeqyrfgqtqflgnlttqaeleallkfaegfsg 304
 Qy 305 AMLEOTTNNISTKFGDDYYIAQIRPVTRINDESRTVDVEYIDPVHPVYRRINFTGNF 364
 Db 305 amleottnnistkfgddyyiaqirpvrtrindestrtdvveyidpvhpyrrinfntgnf 364
 Qy 365 KTQDEVLRMRQLEGALASNQIKQLSRARLMRTGFFKHHVTVTRPVNPSPDQVDV 424
 Db 365 ktqdevlrmerqlegalasnqiklsrarlmrtgffkhhvtvtrpvnpdpqvdv 424
 Qy 425 EQPSGSTTAAGYSQSGVTFQFDVSONNFMCTGKHVNASFSESTREYVSLGNTNYPF 484
 Db 425 eqpsgsttaagysqsgvtfqfdvsonnfmctgkhvnasfsestreyvslgntnpyf 484
 Qy 485 TVNGYSQSLGYRKTXYDNKNISNYVLDYSGGSLSYGYPIDENQRIISFGLNADNTKLHG 544
 Db 485 tvngysqslgyrktxydnknisnyvldysggslsygypidenqrifsglnadntklhg 544
 Qy 545 GRFMGINSVKOLMADGGKIQVDNNGIPDPKHDTYTYNAILGWNYSLSLRPVFTQGMSHS 604
 Db 545 grfmginvskolmadggkiquvdnngipdpkhdtytyynailgwnyslsldrvpftqgmshs 604
 Qy 605 VDLTVGFGDKTHOKVYVQGNIRYRPFIKKSVLRGVAKLGYGNLPPFENYVAGGYSGVRG 664
 Db 605 vdltvvgfkdthokvyvqgniryrfikksvlrgvakiigygnlppfenyvaggysgvrg 664
 Qy 665 DQSSLGPRSOAYLTARRGQTTLGEVYVGGNALTFFGSELILPLPFKGDWIDQVRPV 724
 Db 665 dqsslgprsqayltarrgqttlgevvyggnaltffgselellplpfkgdwidqvrpv 724
 Qy 725 GGOVFTTGMKDQIFDLTQFKDPOATAEQNAKANRPLLTQDKOLRYSGAGVATWTPIG 784
 Db 725 ggovfttgmkdqifdltdqfkdpqataeqnakanrplltqdkqlrysagvgatwtpig 784
 Qy 785 PLISISYAKPLNKKONDQTDVTFQIGSVF 813
 Db 785 plisisyakplnkkondqtdvtfqigsvf 813

AA44393
ID AAY44393 standard; peptide; 17 AA.

XX AC AAY44393;
XX DT 14-MAR-2000 (first entry)
XX DE Moraxella catarrhalis BASB027-specific synthetic peptide #2.
XX KW BASB027; OMP85; outer membrane protein; otitis media;
XX KW bacterial infection.
XX OS Synthetic.

XX PN WO9963093-A2.

XX PD 09-DEC-1999.

XX PF 31-MAY-1999; 99WO-EP03822.

XX PR 03-JUN-1998; 98GB-0011945.
XX PR 08-MAR-1999; 99GB-0005304.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Vinals-Bassois C;

XX DR WPI; 2000-105700/09.

XX PT Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media -
XX Example 8; Page 64; 109pp; English.

XX CC The present sequence is a Moraxella catarrhalis BASB027
CC polypeptide-specific peptide which can be coupled to KLH and used for
CC producing antibodies against BASB027. Anti-BASB027 antibodies can be used
CC to treat bacterial infections.

XX SQ Sequence 17 AA;

Query Match 2.0%; Score 16; DB 21; Length 17;
Best Local Similarity 100.0%; Pred. No. 1e-08;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 676 YLTARRGQQTTLGEVV 591

Db 1 yltarrgqgttlgevv 16

RESULT 4

AA44392

ID AAY44392 standard; peptide; 16 AA.

XX AC AAY44392;

XX DT 14-MAR-2000 (first entry)

XX DE Moraxella catarrhalis BASB027-specific synthetic peptide #1.

XX KW BASB027; OMP85; outer membrane protein; otitis media;
XX KW bacterial infection.

XX OS Synthetic.

XX PN WO9963093-A2.

XX PD 09-DEC-1999.

XX PF 31-MAY-1999; 99WO-EP03822.

XX

PR 03-JUN-1998; 98GB-0011945.
PR 08-MAR-1999; 99GB-0005304.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Vinals-Bassois C;

XX DR WPI; 2000-105700/09.

XX PT Novel BASB027 polynucleotide and polypeptides from Moraxella
PT catarrhalis useful for treating M. catarrhalis infection such as otitis
PT media -
XX Example 8; Page 64; 109pp; English.

XX CC The present sequence is a Moraxella catarrhalis BASB027
XX CC polypeptide-specific peptide which can be coupled to KLH and used for
XX CC producing antibodies against BASB027. Anti-BASB027 antibodies can be used
XX CC to treat bacterial infections.

XX SQ Sequence 16 AA;

Query Match 1.8%; Score 15; DB 21; Length 16;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 790 YAKPINKKQNDQDT 804

Db 2 yakpinkkqndqdt 16

Search completed: September 5, 2002, 10:04:38

Job time: 511 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:05:08 ; Search time 25.36 Seconds
(without alignments)
783.044 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTMAVM.....LNKKONDQDTVTQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length	ID Description

No matches found

Search completed: September 5, 2002, 10:05:08
Job time: 316 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:05:54 ; Search time 39.77 Seconds
(without alignments)
1964.310 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTMAVMM.....LNKKQNDQDTVQFOIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length	DB ID Description

No matches found

Search completed: September 5, 2002, 10:05:54
Job time: 167 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:12:48 ; Search time 20.98 Seconds
(without alignments)
1500.429 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYFKGFQVSAMTMAVM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description

No matches found

Search completed: September 5, 2002, 10:12:48
Job time: 411 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 5, 2002, 10:12:20 ; Search time 63.51 Seconds
(without alignments)
2214.531 Million cell updates/sec

Title: US-09-701-711-4
Perfect score: 813
Sequence: 1 MRNSYKFGQVSAMTAVMM.....LNKKQNDQDTVQFQIGSVF 813

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

- Database : SPTREMBL19.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archheap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length DB ID	Description

No matches found

Search completed: September 5, 2002, 10:12:20
Job time: 428 sec

7.04
2.02
2.4

39-19

June

OM of: US-09-701-711-2 to: EST:* out_format : pfs

Date: Sep 19, 2002 1:01 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=frame+2n.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09701711/runat_17092002_141401_26878/app_query.fasta_1.882
-DB=EST -OFMT=fastap -SUFFIX=sepl7std.rst -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09701711 -CGN1_1_5751 -NCPU=6 -ICPU=3 -LONGLOG..
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-701-711-2

Query length: 813

Database: EST:*

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 1941.050000

score_list:	Sequence	Strd	Orig	zScore	Escore	Len	Documentation
	gb_est2:BF942509	-	253.50	469.87	6.3e-17	620	1
	gb_gss:BH396320	-	235.00	433.49	6.7e-15	649	1
	gb_rtc:AK003990	-	204.00	362.35	6.2e-11	1629	1
	gb_gss:AQ990340	-	190.00	346.57	4.7e-10	636	1
	gb_gss:AQ990755	+	182.50	334.35	2.2e-09	525	1
	gb_est1:AW947726	+	179.00	335.23	2.0e-09	278	1
	gb_gss:BH371368	-	150.50	267.66	1.2e-05	776	1
	gb_est2:BI903558	-	139.50	246.50	0.0002	767	1
	gb_gss:BH369516	+	137.00	244.32	0.0002	615	1
	gb_est2:BE678332	-	136.50	244.76	0.0002	547	1
	gb_est2:BE678801	-	134.50	240.75	0.0004	553	1
	gb_est1:AW115556	+	134.50	238.18	0.0005	685	1
	gb_est1:AA474343	+	134.00	248.88	0.0001	260	1
	gb_est2:BM119288	-	134.00	237.00	0.0006	697	1
	gb_est2:BG303393	-	133.50	236.42	0.0006	675	1
	gb_est2:BG344329	+	133.50	234.31	0.0008	804	1
	gb_est2:BI278122	-	132.50	236.45	0.0006	573	1
	gb_est2:BJ054028	-	132.00	232.72	0.0010	721	1
	gb_est2:BJ088421	-	131.50	236.62	0.0006	481	1
	gb_est2:BM233570	-	131.50	234.37	0.0008	580	1
	gb_est1:BB201163	+	131.00	232.20	0.0011	641	1
	gb_est2:BI539548	+	130.50	233.12	0.0010	548	1
	gb_est2:BI714058	+	130.50	232.64	0.0010	570	1
	gb_est1:AI326111	-	130.50	232.50	0.0011	577	1
	gb_est2:BM205401	-	130.50	232.33	0.0011	585	1
	gb_est1:AI553725	+	130.50	232.21	0.0011	591	1
	gb_est2:BG760879	+	130.50	232.06	0.0011	598	1
	gb_est2:BM293469	-	130.50	231.84	0.0011	609	1
	gb_est1:AV266467	+	130.50	231.47	0.0012	628	1
	gb_est1:BB471006	+	130.50	230.86	0.0013	661	1
	gb_est2:BE612912	-	130.50	229.19	0.0016	759	1
	gb_gss:BH391826	-	130.50	228.74	0.0017	788	1
	gb_est1:AW107426	-	130.50	227.53	0.0020	871	1
	gb_est2:BM233588	-	129.50	230.50	0.0014	580	1
	gb_est2:BI277013	-	125.50	223.74	0.0032	534	1
	gb_est2:BE647176	+	124.50	221.65	0.0042	541	1
	gb_est1:AW027570	-	124.50	220.48	0.0049	596	1
	gb_est1:AW665916	-	123.50	218.51	0.0063	598	1
	gb_est1:AW162459	-	122.50	216.49	0.0082	602	1
	gb_est1:AW150700	-	120.50	213.57	0.0119	556	1

gb_est1:AV611041 - 119.50 212.17 0.0143 532 ! AV611041 AV611041 Bos taurus
gb_est2:BE302194 - 119.50 211.94 0.0147 542 ! BE302194 BB82f11.x1 NIH_MGC
gb_est2:BF983300 + 119.50 210.11 0.0186 631 ! BF983300 60230554F1 NIH_MGC
gb_est2:BF690074 - 119.50 204.37 0.0389 1016 ! BF690074 602186378T1 NIH_MGC
gb_est1:AI386226 + 119.00 214.19 0.0110 415 ! AI386226 mq97c11.y1 Soares m

seq_name: gb_est2:BF942509

seq_documentation_block:

LOCUS BF942509 620 bp mRNA linear EST 22-JAN-2001
DEFINITION LD2 cDNA library of 4-day-old Eucalyptus globulus
bicostata-Pisolithus tinctorius ectomycorrhiza Pisolithus
tinctorius cDNA 5' similar to outer membrane protein, mRNA
sequence.
ACCESSION BF942509
VERSION BF942509.1 GI:12359829
KEYWORDS EST.
SOURCE Pisolithus tinctorius.
ORGANISM Pisolithus tinctorius
Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
Boletales; Sclerodermataceae; Pisolithus.
1 (bases 1 to 620)
Voiblet,C., Duplessis,S., Encelet,N. and Martin,F.
Identification of symbiosis-regulated genes in Eucalyptus
globulus-Pisolithus tinctorius ectomycorrhiza by differential
hybridization of arrayed cDNAs
Plant J. 25 (3), 1-12 (2001) In press
Contact: Martin FM
Equipe de Microbiologie Forestiere
Institut National de la Recherche Agronomique
Centre INRA de Nancy, 54280 Champenoux, France
Tel: +33 383 39 40 80
Fax: +33 383 39 40 69
Email: fmartin@nancy.inra.fr
Insert Length: 620 Std Error: 0.00
Seq primer: Capfinder 5' CDS
POLYA-No.

FEATURES
source
1..620
Location/Qualifiers
/organism="Pisolithus tinctorius"
/strain="441"
/db_xref="taxon:37468"
/clone_lib="cDNA library of 4-day-old Eucalyptus globulus
bicostata-Pisolithus tinctorius ectomycorrhiza"
/tissue_type="mantle and Hartig net hyphae"
/dev_stage="symbiotic, 4 days after contact"
/lab_host="Eucalyptus globulus bicostata"
/note="Organ: ectomycorrhiza; Vector: pBluescript; Site_1:
EcoRI. The cDNA library was constructed from 1 ug total
ectomycorrhiza RNA using the SMART cDNA library
Construction kit (#K1051-1, Clontech, Palo Alto, CA, USA)
according to the manufacturer's instructions. cDNAs were
cloned into EcoRI-digested pBluescript."

BASE COUNT 151 a 175 c 167 g 126 t 1 others
ORIGIN

alignment_scores:

Quality: 253.50 Length: 185
Ratio: 2.185 Gaps: 6
Percent Similarity: 62.703 Percent Identity: 35.676

alignment_block:

US-09-701-711-2 x BF942509/rev ..

Align seg 1/1 to reverse of: BF942509 from: 1 to: 620

642 GlyTyrGly.....AsnAsnLeuProPheTyrGluAsnPheTyrAlaG1 656
|||||

620 GGTACGGTTNGACCGCGCTTCCGCTTTCATGAGAACTACTATGCGGG 571
|||||

656 yGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgS 673
|||||

570 TGGTTTCACTCGGTAGTGGTTTCAGAACAGACACCTTGGTTCTTGGTG 521
 673 erGlnAlaTyrLeuThrAlaArg.....GlyGlnGln 684
 ::|||
 520 GT.....ACCCAAGCGGTGGGTGGGTGACCGGTAAACAG 483
 685 ThrThrLeuGlyGlu.....ValValGlyGlyAsnAl 695
 |||::|
 482 GGGACTGTGGTGAATCGGACAGACCGCTGCCATTCGTGGTAAACGT 433
 695 aLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheLeuGly 712
 ::|||
 432 GCTGATCAGGGTGGTGTGAGATTCGTTCGCCGCTGCCATTCGTAAAG 383
 712 spTrpIleAspGlnValArgProValIlePheIleGluGlyGlnVal 728
 |||::|
 382 ATCAA...CGTCCCTCGTACTTCGGTCTTCGGGATGCGGTAAAGTA 336
 729 PheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLe 745
 |||::|
 335 TTCGACTCC.....AA 325
 745 sAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProL 762
 |||::|
 324 GTGGACACAGTCAACACCGGATGGCTCGAGTCCAAACACGAGTGCA 275
 762 euLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThr 778
 ::|||
 274 ACAGCTGTGAGTTCAGCAACCTGGCAAGCTCTGTGGGTGCGGTGAC 225
 779 TrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAs 795
 |||::|
 224 TGGGTACCGCGTGGTCCCATTCAGCTTGTGCTGCGCCATGCGGTCAA 175
 795 nLysLysGlnAspGlnAspGlnValGlnPheGlnIleGlySerV 812
 ::|||
 174 GAACCGGATAACGCTGAACCCAGATTTCCAAATTCCTCCCTCGGCCA 125
 812 alphe 813
 ::|||
 124 CGTTC 120

seq_name: gb_gss:BH396320

seq_documentation_block:

LOCUS BH396320 649 bp DNA linear GSS 11-DEC-2001
 DEFINITION AG-ND-161M3.TR ND-TAM Anopheles gambiae genomic clone AG-ND-161M3,
 DNA sequence.

ACCESSION BH396320.1 GI:17342461

VERSION GSS.

KEYWORDS

SOURCE African malaria mosquito.

ORGANISM

Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
 ; Anopheles.

REFERENCE

1. (bases 1 to 649)
 Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

Direct Submission of BAC-end sequences from Anopheles gambiae

Unpublished (2001)

Other GSSs: AG-ND-161M3.TF

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit

the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.
 Seq primer: M13 Rev
 Class: BAC ends.

FEATURES

source
 1. .649
 /organism="Anopheles gambiae"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone="AG-ND-161M3"
 /clone_lib="ND-TAM"
 /note="Vector: pECBAC1; Site_1: HindIII"
 BASE COUNT 146 a 182 c 183 g 138 t
 ORIGIN

alignment_scores:

Quality: 235.00 Length: 178
 Ratio: 2.156 Gaps: 6
 Percent Similarity: 61.236 Percent Identity: 32.022

alignment_block:

US-09-701-711-2 x BH396320/rev ..

Align seg 1/1 to reverse of: BH396320 from: 1 to: 649

637 GlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyrGluAsnPh 653
 |||::|
 557 GGTGGGCAAGGCCCTGGGGCGCAACCCCTATCCCATCTCAAGAACTT 508
 653 eTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuG 670
 |||::|
 507 CTATTCGGCGGCTGCTGGTTCGGTACCGGTTTCGACGAGGGCTCCCTGG 458
 670 lyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnGlnThr 686
 |||::|
 457 GCAAGCACGATTCGGTG.....AACACCAAC 432
 687 LeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLe 703
 |||::|
 431 CTG.....GCCCTCGCGGTACACGCAAGCTGACCTGAATGGCGAGTT 388
 703 uileLeuProLeuProPheLysGlyAspTrpIleAspGln...ValArgp 719
 ::|||
 387 CATGGTTCATTCCTCCCGGAGCCGCAAC.....GACAGAACCTTGGCTC 344
 719 roValIlePheleGluGlyGlyGlnValPheAspThrThrGlyMetasp 735
 |||::|
 343 TGTTCGGCTTCGTGGACGTGGTAATGTGG.....GCGGAA 306
 736 LysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGl 752
 ::|||
 305 GGTGAAGACATGATCTG..... 288
 752 uGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLysGlnL 769
 287GGGACCT 281
 769 euArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleGlyPro 785
 |||::|
 280 TGGGTGCTTCACCGGTATCGGTATCAGCTGGATTTTCGGCGCTGGGTCG 231
 786 LeuSerIleSerTyrAlaLysProLeuAsnLysGlnAsnAspGlnTh 802
 |||::|
 230 TTGGTTTGGCGTATCGCTACCTATCCGTAAGGAACACCGCGGATAGAT 181
 802 rAspThrValGlnPheGlnIleGlySerValPhe 813
 ::|||
 180 CCAGACCATCAATTCCAATCGGAACATCTTTC 147

seq_name: gb_htc:AK003990

```

seq_documentation_block:
LOCUS       AK003990               1629 bp      mRNA       linear      HTC 19-JAN-2002
DEFINITION Mus musculus 18 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:1110030L07:homolog to PROTEIN CGI-51, full
            insert sequence.
ACCESSION   AK003990
VERSION     AK003990.1 GI:112834981
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) 18 days embryo cDNA to mRNA,
            clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:1110030L07.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (sites)
AUTHORS     Carninci,P. and Hayashizaki,Y.
TITLE       High-efficiency full-length cDNA cloning
JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
MEDLINE     99279253
PUBMED      10349636
REFERENCE   2 (sites)
AUTHORS     Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL     Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE     20493374
PUBMED      11042159
REFERENCE   3 (sites)
AUTHORS     Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M.,
            Tameda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE       RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE     20530913
PUBMED      11076861
REFERENCE   4 (sites)
AUTHORS     The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE       Functional annotation of a full-length mouse cDNA collection
JOURNAL     Nature 409, 685-690 (2001)
MEDLINE     11076861
PUBMED      11076861
REFERENCE   5 (bases 1 to 1629)
AUTHORS     Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurlhar,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ono,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome

```

Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGCGCGCACTGAGTCTTTTATTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAGATCCAGAGCTCAATTAATTAATTAACCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

FEATURES

source	1..1629	Location/Qualifiers
	/organism="Mus musculus"	
	/strain="C57BL/6J"	
	/db_xref="MGI:1900625"	
	/db_xref="taxon:10090"	
	/clone="1110030L07"	
	/clone_lib="RIKEN full-length enriched mouse cDNA library"	
	/dev_stage="18 days embryo"	
misc_feature	1..1629	
	/note="data source:SPTR, source key:09Y512, evidence:ISS homolog to PROTEIN CGI-51"	
BASE COUNT	376 a 392 c 484 g 377 t	
ORIGIN		

alignment_scores:

Quality:	204.00	Length:	521
Ratio:	0.761	Gaps:	24
Percent Similarity:	51.440	Percent Identity:	22.265

alignment_block:

US-09-701-711-2 x AK003990 ..

Align seg 1/1 to: AK003990 from: 1 to: 1629

```

334   ileAsnAspGluSerArgThrValAspValGlu.....Tyrty 346
151   CTGGGGGAGGAGCGGAGTTTGTGGNAGTTGAGCGGAGCGGAGGACGGA 200
346   rleAspProValHisProValTyrValArgArgIleAsnPheThrGlyA 363
201   ATCTCTGGAACCAAGATGTGTGTGTCAGCATGTTTCATTGATGGAC 250
363   snPheLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeu... 378
251   TTGGCGGAGCTAAGGATGACATCATCTTTGTGAAATCGGAGAGGTCCTT 300
379   .....GluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAl 393
301   RAGGCTAAAAACCTCATTCAGGTAAATCGCGCATCTCATGACCGCGGA 350
393   aArgLeuMetArgThrGlyPhePheLysHisValThrVal.....Aspt 408
351   AAAACTGTTCTCGCTAGGAATTTTAGACACAGTGGATGTTTTCATCGATA 400
408   hrArgProValProAsnSer....ProAspGlnValAspValAsnPheVal 423
401   CATGTCATGGTGAAGATGCCCTGCCCAATGGGTGTAGATGTCACCTTTGAA 450
424   ValGluGlu.....GlnProSerGlySerSerThrIleAlaAlaGlyTy 438
451   CTGACAGAGCTCAGGAGACTGACGGCGAGTTACACACCATGTTGGAAA 500
438   rSerGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnP 455
501   CAACGAAGGCGAGTATGGTA.....CTCGGCCTCAAACTCCCAACC 541
455   heMetGlyThrGlyLysHisValAsnAlaSerPheSerArgSerGluThr 471
542   TTCTGGGACGACGAGAAAAAGTCACTTTCCAGTTTTTCTTATGGAACCAA 591

```

```
472 ArgGluValTyrSerLeuGlyMetThrAsnProTyr..... 483
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 GAAACCTTCAGCGCTGCTCTTCTCAAGCCACACGCTGGAACCTCGA 641
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
484 .....PheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyr. 497
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
642 GAGAAATTTCTCGTAAC.....TTATATAAAGTTACTGGCGCAGTTCC 685
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
498 .....ArgLysThrLysTyrAspAsnLysAsnLysSerAsn 509
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
686 CGTGGAGCTCACTTCGGGAGACA..... 708
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 TyrValLeuAspSerTyrGlyGlySerLeuSerTyrGlyTyrProLeuAs 526
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
709 .....GACAGAGAGTGCTGCAGAGTACAGTTCTCCCTCTGTG 746
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 pGluAsnGlnArg.LleSerPheGlyLeuAsnAlaAspAsnThrLysLeu 542
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
747 GAAGACCAGTCACACTGTCAAGTGG.....AGGGTG 778
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543 HisGlyGlyArgPheMetGlyLleSerAsnValLysGlnLeuMetAlaAs 559
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
779 TGTGGCGGAGCTGGCTGCTCGAGGACTGCTGCTGCTGCTGCTGCGGA 828
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
559 pGlyGlyLysLleGlnValAspAsnAsnGlyLleProAspPheLysHisA 576
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
829 AGGA.....AAGTGA 839
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
576 sPTyrThrThrTyrAsnAlaLleLeuGlyTyrAsnTyr.SerSerLeuAs 592
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
840 CACTCACTGAAGTCGTCTCTCGCATGCCATGCTCATCGACTCTCGAAA 889
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
592 pArgProValPheProThrGlnGlyMetSerHisSerValAspLeuThrV 609
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
890 TTCATCTATCTTGCCAAAGAGAGGGCCCTGTGTTCAAGTCAACGAGGAC 939
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
609 aL...GlyPhe.....GlyAspLysThrHisGlnLysValValTyrGln 622
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
940 TGGCAGGCTACACTGGAGGAGAGTGTGAGCTTCATCAAGGAACTTTGAG 989
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
623 GlyAsnLleTyrArgProPheLleLysLysSerValLeuArgGlyTyrAl 639
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
990 CTTCACTGAATAAGCGCTGCGCTTGGACTCGGTATTCTCC.....AC 1033
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
639 aLysLeuGlyTyrGlyAsnAsnLeuPro.....P 649
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1034 GTCCTCTGGGTGGGAATCGTGTGCCATCGGTGACAGCCATCCAGCA 1083
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
649 heTyrGluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAsp 665
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1084 TTGCTGACAGGTTTACCTGGGAGGCCCCACGAGTGTCCGAGATTAGC 1133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
666 GlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgL 682
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1134 ATGCAGCAGATTGGACCCAGAGTGA..... 1160
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
682 yGlnGlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrP 699
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1161 .....GGAGATTACTGGCGCGGCGGAGGCTACTGGGCTG 1194
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
699 heGlySerGluLeuLleLeuProLeuProPheLys.....GlyAsp 712
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1195 GGGCGCTGCACCTCTACACCCACCTGCCCTTCGGCCAGGCGGAGTGGC 1244
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
713 TrpIleAspGlnValArgProValIlePheIleGluGlyGlyGlnValPh 729
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1245 TTCGGAGAGCTTTTCAGAACTCACTTTTCTCCCAATGCGGCAACCTG.. 1292
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
729 eAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysA 746
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1293 .....TGCACCTCACTATGTTGAGG 1314
```

```
746 sProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeu 762
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1315 GCCCAAAGCCCATATCCGAAGCTAGCTGAGTGC..... 1349
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
763 LeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTr 779
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1350 .....ATCCGCTGCTCTATGAGCAGCGCTGCTCT 1381
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
779 p...TyrThrProLleGlyProLeuSerLleSerTyrAlaLysProLeuA 795
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1382 CGACTTGGCAACATCGCTGGCTGAGCTGAACCTACTGATTCCTATGG 1431
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
795 snLysLysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGly 810
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1432 GCGTGCAGGGGGCGACAGATTGTGTGATGGTGTCCAGTTTGGAGCTGG 1481
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
811 SerValPhe 813
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1482 ATTCGGTTC 1490
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: gb_gss:A0990340
seq_documentation_block:
LOCUS A0990340 636 bp DNA linear GSS 14-AUG-2000
DEFINITION Rf011100 Photorhabdus luminescens strain W14 M13 library
            Photorhabdus luminescens genomic clone PLG01100, DNA sequence.
ACCESSION A0990340
VERSION A0990340.1 GI:9648934
KEYWORDS GSS.
SOURCE Photorhabdus luminescens.
ORGANISM Photorhabdus luminescens.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Photorhabdus
REFERENCE 1 (bases 1 to 636)
AUTHORS ffrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
            Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium
            Photorhabdus luminescens W14: potential implications for virulence
JOURNAL Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE 20378633
COMMENT Contact: ffrench-Constant RH
            Department of Biology and Biochemistry
            University of Bath
            South Building, Bath BA2 7AY, UK
            Tel: (44) 1225 826621
            Fax: (44) 1225 826779
            Email: bsrfe@bath.ac.uk
            This is one of 2,122 random reads from the M13 library. For
            annotation of identified clones (BLASTX, BLASTN and mapping to E.
            coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic
            Acids Res.
            Seq primer: M13 Forward
            Class: Shotgun.
FEATURES             Location/Qualifiers
     source            1..636
                        /organism="Photorhabdus luminescens"
                        /strain="W14"
                        /db_xref="taxon:29488"
                        /clone="PLG01100"
                        /clone_lib="Photorhabdus luminescens strain W14 M13
                        library"
     dev_stage="primary phase variant"
     note="Genomic DNA from strain W14 was size selected (1-2
            kb) and then cloned into M13 Janus."
BASE COUNT 181 a 159 c 131 g 161 t 4 others
ORIGIN
alignment_scores:
Quality: 190.00 Length: 118
Ratio: 2.235 Gaps: 2
Percent Similarity: 72.034 Percent Identity: 36.441
```

/strain="W14"
/db_xref="taxon:29488"
/clone="PLG01576"
/clone_lib="Photorhabdus luminescens strain W14 M13 library"
/dev_stage="primary phase variant"
/note="Genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into MJ3 Janus."

BASE COUNT 160 a 94 c 114 g 154 t 3 others
ORIGIN

alignment_scores:
Quality: 182.50 Length: 178
Ratio: 1.534 Gaps: 4
Percent Similarity: 66.854 Percent Identity: 29.775

alignment_block:
US-09-701-711-2 x AQ990755 ..

Align seg 1/1 to: AQ990755 from: 1 to: 525

196 SerAspAlaAspLeuIleValLeuAlaIleLysAspAsnLysIleAs 212
||||| : ::::: ||||| : : : : : ||||| :
2 TCTTCTCATGAGTTATGAACGTTCCTCCAACTTAGAGATGATGGT 46

212 nPro.....LeuSerLysAlaAspArgTyrThrGlnGluLysLeuV 226
||| : : : : : ||| : : : : : ||| : : : : : ||| :
47 .CCGTGGTGGAATTGACTGCCGATCAGAAATATCAGAAAACAATACTGA 95

226 alThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheVal 242
: : : : : ||||| : : : : : ||||| : : : : : ||||| :
96 CTGTGACCTTGAAGCATTCGCAGGTTTTTACCCTTGATCGCGTTACGCC 145

243 ArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnAr 259
||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| :
146 CGTTTAACATGATTCGACTCAGGTCAGTTGCAGCCAGATAAAAAGG 195

259 gilePheValIleSerLeuHisGluGlyGluInTyrArgPheGlyG 276
||| : : : : : ||| : : : : : ||| : : : : : ||| :
196 TATTATGTCAGCAATAATACTGAAGGGATCAATACAAATATATCAG 245

276 lnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAla 292
: : : : : ||||| : : : : : ||||| : : : : : ||||| :
246 GTATTGACCTGAACGGTAATATGCGGGTTATCAGTCAGAAATTTACTAAA 295

293 LeuLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeuGluGl 309
||| : : : : : ||| : : : : : ||| : : : : : ||| :
296 CTGGCTCTGATGAGCCTGGATCTCTGATACAGGGAGTCAGGTACTAA 345

309 nThrThrAsnAsnIleSerThrLys.PheGlyAspAspGlyTyrTyr 325
: ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| :
346 AATGGAATGACATCAAGAATTTGCTTTGGTCGCTATGGTTATGCTTAC 395

326 AlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrValAs 342
:
396 CCACGAGTGAT.GACGACCTGAAATCAATGATCAAGACAGACAGTNA 444

342 pValGluTyrTyrIleAspProValHisProValTyrValArgArgIleA 359
: : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| :
445 ACTGCATGTCATATTTGATGCGNGGNACCGTTTCTATGTCGTAATAATTC 494

359 snPhe.ThrGlyAsnPheLysThrGlnAsp 368
: : : : : ||||| : : : : : ||||| : : : : : ||||| :
495 GTTCTCTGCTAATGACACCACCTAAAGAT 524

seq_name: gb_est1:AW947726

seq_documentation_block:
LOCUS AW947726
DEFINITION RC0-MT0004-130300-011-c06 MT0004 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW947726
VERSION AW947726.1 GI:8125409

```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
MEDLINE
COMMENT
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-RCO-MT0004-130
300-011-c064t3-2000-03-13&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 61
High quality sequence stop: 278.
FEATURES
source
1. .278
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MT0004"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site:1: SmaI; Site:2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 67 a 82 c 74 g 55 t
ORIGIN
alignment_scores
Quality: 179.00 Length: 90
Ratio: 2.486 Gaps: 4
Percent Similarity: 80.000 Percent Identity: 48.889
alignment_block:
US-09-701-711-2 x AW947726 ..
Align seg 1/1 to: AW947726 from: 1 to: 278
416 AspGlnValAspValAsnPhValGluGlnProSerGlySerSe 432
|||||
17 GACCAAGGTAGATGTGACTAGCGCGTAGAAGAACAAGCCTCAGGTTCCAT 66
|||||
432 rThrIleAlaLaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheA 449
|||||
67 CACGCCAGCGTCGGTAGCGTAGAGATGCGGTTGTATCCTGGGTGGTT 116
|||||
449 spValSerGlnAsnAsnPhMetGly.ThrGlyLysHis.ValAsnAlas 465
|||||
117 CCATCAGCAGACAACATCTCTGGGATACCGGTAACCGATGTGTCATCG 166
|||||
465 erPheSerArgSerGlu.ThrArgGluValTyrSerLeuGlyMetThrAs 481
|||||
167 GCCTGACCCGAGCGGAATACAGAGACCGACCAACTTCGCGTATACCGA 216
|||||
481 nProTyrPheThrValAsnGlyValSerGlnSerLeu.SerGlyTyrTyr 497
|||||
217 CCCCTACTCGACTGCAGCGGTGAGCGCTGGTGCTACTACGCCCTTCTAC 266
|||||
498 ArgLysThrLys 501
|||||
267 CGCACCACCGCAA 278
|||||
seq_name: gb_gss:BH371368
seq_documentation_block:
LOCUS BH371368 776 bp DNA linear GSS 10-DEC-2001
DEFINITION AG-ND-175A22.TR ND-TAM Anopheles gambiae genomic clone AG-ND-175A22
, DNA sequence.
ACCESSION BH371368
VERSION BH371368.1 GI:17317493
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
; Anopheles.
REFERENCE
1 (bases 1 to 776)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-175A22.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 Rev
Class: BAC ends.
FEATURES
source
1. .776
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone_lib="AG-ND-175A22"
/note="Vector: pECBAC1; Site:1: HindIII"
BASE COUNT 213 a 140 c 120 g 303 t
ORIGIN
alignment_scores
Quality: 150.50 Length: 240
Ratio: 1.060 Gaps: 9
Percent Similarity: 59.167 Percent Identity: 22.500
alignment_block:
US-09-701-711-2 x BH371368/rev ..
Align seg 1/1 to reverse of: BH371368 from: 1 to: 776
65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPhSerAs 81
|||||
698 AGACTGAGTACAGCTATCAAAAGCTTTGGGATAGCCCAATTTTCTGA 649
|||||
81 pValGlnValTyr.....HisGlnGluGlyArgIleI 92
|||||

```



```

:||||:|||||
648 AGTTGAAGTATATGTCAGACATTATAGCACACATTAAAGGCACAAATG 599
92 leTyrGlnValThrGlu.....ArgProLeuIleAlaGluIleAsn 105
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
598 TCTTAATCTCTCTTCAGGATTTAAAGAACTTGGTAAGTAATAATTC 549
106 PheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 122
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
548 GGAAGAAGCA.....ATCAAAATCTTAAACGAAACAACTTATTAA 508
122 sasnaAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGln 139
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507 GGATAATAACCTGAAGCCTGTGTATGAATAATACCGAAGACCTGGTTACTA 458
139 etIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsn 155
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
457 ACCTAAAGCAATAGTACTCAGCAGATACATCAAAAGGATTCCTGGAT 408
156 ThrGluIleThrVal.....LysGlnThrMetLeuAspGlyAsnArgVa 170
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
407 GCTAAGATCACCATTGAAGATATAAATCAATGCAAAAGGATCTAGCTTAT 358
170 lLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspI 187
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
357 CGACTGGACCATTAATGTAGATAAAGGAAACGTTTAAAGATAGACCGTA 308
187 leAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu.Ile...As 202
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
307 TAGATTTGAAGGAACAAACAGTGTCTTCATCAAACTTCGTAAAC 258
202 pValLeuAlaIleLysAspAsnLys.....IleAsnProLeuSerLysA 217
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257 GGTTTTAAATACGAAACAAAGAGATCTCTCTGTTGCTTAAGC 208
217 laAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArg 233
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
207 CTTCGAATTTATCAAGATAAATATGAAGAGATGAAGAAGACTCTGTA 158
234 AlalysTyrLeuAsnAlaGlyValArgPheGluIleLysAspAlaLys 250
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
157 GACTACTACAACCTTTAGGATTT.....AGAGATATAG 123
250 sLeu.....AsnIleAsnGluAspLysAsnArgIlePheValGluI 264
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
122 AGTAGTGCAGACTCTGTTTCGCTTAATACTTAAGGATACAACATCAAG 73
264 leSerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeu 280
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
72 TAAAGTAGATAGAGGGTAAAAAATACTATATCGGAGATATTACATTCGA 23
281 GlyAsnLeuThrTyrThr 286
||||| : : : : :
22 GGAATACAGATTATTTCT 5
```

seq_name: gb_est2:BI903558

```

seq_documentation_block:
LOCUS      BI903558                      767 bp    mRNA    linear    EST 16-OCT-2001
DEFINITION 603168388F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5256452 5',
            mRNA sequence.
ACCESSION  BI903558
VERSION    BI903558.1  GI:16165506
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 767)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
```

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM11647 row: p column: 21
High quality sequence start: 22
High quality sequence stop: 764.
Location/Qualifiers
1. .767
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5256452"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT730-Pac (Pharmacia) with a modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCTCTGTTTTTTTTTTTTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT730 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES

source

BASE COUNT 175 a 199 c 221 g 171 t 1 others
ORIGIN

alignment_scores:

Quality: 139.50 Length: 239
Ratio: 1.107 Gaps: 10
Percent Similarity: 52.720 Percent Identity: 25.105

alignment_block:

US-09-701-711-2 x BI903558 ..

Align seg 1/1 to: BI903558 from: 1 to: 767

```

590 SerLeuAspArgProValPheProThrGlnGlyMetSerHisSerValas 606
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3 ACTCTCAACCCGCTGCTGTTATTCGAAGAAGAGGGGCTTGTTCAAAGTCAA 52
606 pLeuThrVal...GlyPhe.....GlyAspLysThrHisGlnLysValV 620
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| |||
53 CCAGGAGCTGGCAGGCTACACTGGAGAGATGTGAGCTTCATCAAGGAAG 102
620 alTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeuArg 636
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
103 ACTTTGAGCTTCAGCTGAATAGCGCTCGCTTGGACTCGGTATTCTCC 152
637 GlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuPro..... 648
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 .....ACGTCTCTGGGGTGGAAATGCTGGTGGCCCATCGGTGACAAAGCC 196
649 .....PheTyrGluAsnPheTyrAlaGlyTyrGlySerValArgG 663
: : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 ATCCAGCATTCGTGACAGGTTTACCTGGAGGCCCCAGAGCTGCTCCGAG 246
663 lyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAla 679
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
247 GATTAGCATGCACAGCATTTGAGCCCAAGAGTAA..... 281
680 ArgArgGlyGlnGlnThrLeuGlyGluValValGlyGlyAsnAlaLe 696
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282 .....GGAGATTACCTGGCGCGCGAGGCCTA 307
```

```

696 uAlaThrPheGlySerGluLeuLeuProLeuProPheLys..... 710
   ::  |||  ::|||  |||||  |||||  ::
308 CNGGGCTGGGGCTGCACCTACACCCACTGCCCTCCGGCCAGGCC 357
711 ..GlyAspTrpIleAspGlnValArgProValIlePheIleGluGly 726
   |||  ::  ::  |||  ::|||  ::|||  ::|||  ::|||  ::|||
358 AGGGTGGCTTCGGAGAGCTTTTCAGAACTCACTTTTCTCAATCGGGC 407
727 GlnValPheAspThrGlyMetAspLysGlnThrIleAspLeuThrG1 743
   ::|||  ::|||  ::|||  ::|||  ::
408 AACCTG.....TGCACCTCACTA 427
743 nPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaLysnA 760
   ::  |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
428 TGGTGAGGCCCAAGCCCATATCCGAAGCTAGCTGAGTGC..... 470
760 rgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGly 776
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
471 .....ATCCGCTGGTCTATGGAGCAGGC 494
777 AlaThrTrp...TyrThrProIleGlyProLeuSerIleSerTyrAlaLys 792
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
495 GTCGCTCCGACTGGCAACATCGCTCGGCTGGAGCTGAACCTACTGCAT 544
792 sProLeuAsnLysLysGlnAsnAspGln...ThrAspThrValGlnPheG 808
   |||||  ::  ::  |||||  ::  ::  |||||  ::  ::  |||||
545 TCCTATGGCGTCGACGGGGGACAGGATTNGTGATGGTGTCCTCAGTTG 594
808 InIleGlySerValPhe 813
   |||  |||
595 GAGCTGGGATTCGGTTC 611

```

seq_name: gb_gss:BH369516

seq_documentation_block:

LOCUS BH369516 615 bp DNA linear GSS 10-DEC-2001
 DEFINITION AG-ND-150J12.TF ND-TAM Anopheles gambiae genomic clone AG-ND-150J12

ACCESSION BH369516

VERSION BH369516.1 GI:17315617

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
 ; Anopheles.

REFERENCE 1 (bases 1 to 615)

Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
 Direct Submission of BAC-end sequences from Anopheles gambiae
 Unpublished (2001)

Other_GSSs: AG-ND-150J12.TR

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

tel: 301 838 0208

Fax: 301 838 3543

Email: bjoftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by
 F.H. Collins and sequenced by The Institute for Genomic Research
 (TIGR). The BAC library was generated from A. gambiae PEST strain
 DNA. All DNA was extracted from newly hatched first instar larvae
 to minimize the inclusion of DNA from microorganisms that inhabit
 the gut. The DNA is derived from mixed sexes of larvae. The BAC
 library was constructed at Texas A&M University BAC Center
 University, College Station, Texas 77843-2123, USA using a HindIII
 partial digest.

Seq primer: M13 For

Class: BAC ends.

Location/Qualifiers

1..615

/organism="Anopheles gambiae"

/strain="PEST"

FEATURES

source

```

/db_xref="taxon:7165"
/clone="AG-ND-150J12"
/clone_lib="ND-TAM"
/notes="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT 198 a 114 c 132 g 171 t
ORIGIN

```

alignment_scores:

Quality: 137.00 Length: 227

Ratio: 1.062 Gaps: 9

Percent Similarity: 56.828 Percent Identity: 24.229

alignment_block:

US-09-701-711-2 x BH369516 ..

Align seg 1/1 to: BH369516 from: 1 to: 615

```

277 ThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLe 293
   |||  ::|||  |||||  |||||  ::
8 ACAAAACCTCTTGGT.....TACAAAAAGGTGATATTACGATTC 48
293 uLeuLysPheLys.....AlaGluGluGlyPheSerGlnAlaMetL 307
   ::  |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
49 CGTAGTTTCAGAAAAAAGTAGGTGAAGAGGTGGT..... 85
307 euGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyr 323
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
86 ..AAAGAAGATAACTCTGATATCGCTTCATCTTATATGATAGTAGTGTAC 133
324 TyrTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThr 340
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
134 CTTTCTCTAATGTAATGCAGTGGAGAAA...TCTATTAAAGACGATAC 180
340 rValAspValGluTyrTyrIleAspProValHisProValTyrValArgA 357
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
181 CATCAGTAGGAAAGTTCGTATCCAGAGGTACAAAACGACATGGATTC 230
357 rgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArg 373
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
231 GTGTAACATGGGGCGGAACGTTACTACACATGACCATGTTATCTTAAGA 280
374 GluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLe 390
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
281 TCATTAAAGAACACGCTCGGAGATCTTCTTAAGGCAACATTAAGAAG 330
390 userArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValA 407
   ::  |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
331 AACTTATTTCGATCTTCGGGGATGCTTACTTCGCCCTCAGCAATCG 380
407 spThrArgProValProAsnSerProAspGln...ValAspValAsnPhe 422
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
381 GACAGGATATTAAAGCTAATGCAGTAGACAATACAGCAGATATTCAGTGG 430
423 ValValGluGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSe 439
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
431 ACTGTAGTAGAAAAAGGATCTTCTCAGGTTTCAGGTCAGGCGAGGTAT.. 478
439 rGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheM 456
   |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
479 .....GGTGGT.....AACTCTTTCA 494
456 etGlyThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArg 472
   ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
495 TCGGAACC.....TTAGTCTTACTTTCAACAACCTCTCTCTGAA 535
473 GluValTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyVa 489
   ::  |||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
536 AATTTCCTCAGATTGAAAGACTTTCAGCA.....GTTCTCTCAGGTGA 579
489 lserGlnSerLeuSerGlyTyrTyrArgLys 499
   ::|||  |||||  ::|||  ::|||  ::|||  ::|||  ::|||  ::|||
580 CGGACAAATCCTTTCTTACAGGCACAGGCAG 610

```

seq_name: gb_est2:BE678332

seq_documentation_block:

LOCUS BE678332 547 bp mRNA linear EST 18-APR-2001
 DEFINITION df54c02.x1 xenopus laevis unfertilized egg cDNA library Xenopus
 laevis cDNA clone IMAGE:3743138 3' similar to SW:YDF6_SCHPO Q10478
 HYPOTHETICAL 51.8 KD PROTEIN C17C9.06 IN CHROMOSOME I.; mRNA
 sequence.

ACCESSION BE678332
 VERSION BE678332.1 GI:10059901

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 547)

AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
 Martin,J., Wyllie,T., Underwood,K., Theising,B., Bowers,Y., Person
 , B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
 Waterston,R. and Wilson,R.

TITLE WashU Xenopus EST project, 1999

JOURNAL Unpublished (1999)

COMMENT

Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library constructed by Bruce Blumberg

Library normalized by Jiwhan Song

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: Xenopus clone distribution information for

this library can be found through Research Genetics, visit their

web page at: <http://www.resgen.com/> Please reference the id listed

below when ordering this clone: Source lab clone id - xlnneg002b03

Seq primer: -40UP from Gibco

High quality sequence stop: 477.

FEATURES

source

1..547

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone_lib="Xenopus laevis unfertilized egg cDNA library"

/tissue_type="unfertilized egg"

/lab_host="Top-10 F"

/note="Vector: pBluescript SK-; Site1: EcoRI; Site2:

XhoI; cDNA was prepared from 2ug of poly A+ RNA.

EcoRI-XhoI cut cDNA was then ligated into Unizap-XR

(Stratagene) with EcoRI at the 5' end and XhoI at the 3'

end. SS-library phagemids were prepared by mass excision

from the original library and normalized by hybridization

to biotinylated driver (prepared from the same library by

PCR) to Cot-omega of 11. After removal of hybrids and

excess driver by streptavidin sepharose chromatography,

the ss-phagemids were made double stranded and

electroporated into Top-10 F'. Original library

contruction by Bruce Blumberg (Blumberg et al., 1991

Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,

2923-2935). Note: This is a Xenopus Gene Collection (XGC)

library."

BASE COUNT 131 a 161 c 136 g 119 t

ORIGIN

alignment_scores:

Quality: 136.50 Length: 168

Ratio: 1.517 Gaps: 6

Percent Similarity: 53.571 Percent Identity: 27.381

alignment_block:

US-09-701-711-2 x BE678332/rev ..

Align seg 1/1 to reverse of: BE678332 from: 1 to: 547

```

651 GluAsnPhetYrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      :::::|||||  |||||  |||||  |||||  |||||  |||||  :::::
544 GACGGTGTTCCTGGGGGACCCACGAGCGTGGCGAGGATTGAGCATGTA 495
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
494 CAGCACTGGCCCCCAGACGAA..... 473

684 lnThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
472 .....GGCGACTATTGGCGCGAGAACTTACTGGGCTGGGGC 434
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
701 SerGluLeuLeuLeuLeuProPhe.....LysGlyAspTrpIl 714
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
433 GTGCACTTATACACCCCTCTACCCCTTCGCGCGGACGCGGGGATTGG 384
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
714 eAspGlnValArgProValIlePheIleGluGlyGlnValPheAspT 731
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
383 GGACCTTTTCAGACCCCACTTCTCTCAACGCTGGGAACCTC..... 341
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
340 .....TGCAACCTGAACCTACGCGGAGGGGCCCC 314
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
313 AGGCCCATCTCGACGCGATTGGCTGAATGT..... 284
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
283 .....ATCCGCTGGTCTATGCGGGGGGCTAGTTCGCGCT 247
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
246 TGGGGAATATCGCAGACGTGGAGCTGAATTATTGTATCCCTATGGGGTC 197
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
      :::::  |||||  |||||  |||||  |||||  |||||  :::::
196 CAGAGTGGAGAGAGGATCTGCGATGGGTTTCAGTTCGCGCTGGAATCCG 147
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
812 lphe 813
      |||||
146 ATTC 143

```

seq_name: gb_est2:BE678801

seq_documentation_block:

LOCUS BE678801

DEFINITION df76902.x1 Xenopus laevis oocyte non normalized Xenopus laevis cDNA
 clone IMAGE:3745298 3' similar to SW:YNC6_YEAST P53969 HYPOTHETICAL
 54.4 KD PROTEIN IN KTR5-UME3 INTERGENIC REGION.; mRNA sequence.

ACCESSION BE678801

VERSION BE678801.1 GI:10060840

KEYWORDS EST.

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 553)

AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
 Martin,J., Wyllie,T., Underwood,K., Theising,B., Bowers,Y., Person
 , B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
 Waterston,R. and Wilson,R.

TITLE WashU Xenopus EST project, 1999

JOURNAL Unpublished (1999)

COMMENT Contact: Sandy Clifton, Ph.D.

WashU Xenopus EST project, 1999

Washington University School of Medicine

```

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jiwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: http://www.resgen.com/ Please reference the id listed
below when ordering this clone: Source lab clone id - xlnnoc001103
Seq primer: -40UP from Gibco
High quality sequence stop: 476.
Location/Qualifiers
    1..553
      /organism="Xenopus laevis"
      /db_xref="taxon:8355"
      /clone="IMAGE:3745298"
      /clone_lib="Xenopus laevis oocyte non normalized"
      /tissue_type="oocyte (stages 5 and 6)"
      /lab_host="Top-10 F'"
      /note="vector; pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA.
EcoRI-XhoI cut cDNA was then ligated into Unizap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'
end. SS-library phagemids were prepared by mass excision
from the original library and normalized by hybridization
to biotinylated driver (prepared from the same library by
PCR) to Cot-omega of 11. After removal of hybrids and
excess driver by streptavidin sepharose chromatography,
the ss-phagemids were made double stranded and
electroporated into Top-10 F'. Original library
construction by Bruce Blumberg (Blumberg et al., 1991
Science 253, 194-196; Hawlet et al., 1995, Genes Dev. 9,
2923-2935). Note: This is a Xenopus Gene Collection (XGC)
library."
BASE COUNT      132 a   161 c   139 g   121 t
ORIGIN

alignment_scores:
    Quality: 134.50      Length: 168
    Ratio: 1.494        Gaps: 6
    Percent Similarity: 53.571    Percent Identity: 27.381

alignment_block:
US-09-701-711-2 x BE678801/rev ..

Align seg 1/1 to reverse of: BE678801 from: 1 to: 553

651 GluasnPhetYrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| :
547 GACCGGTTTACTGGGGGGACCAGCAGCGTGCGAGGATTACCATGTA 498
|||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| :
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| :
497 CAGCATGGCCCCCAGAGCGAA..... 476

684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
|||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| :
475 .....GGAGACTATTTGGGTGGAGAAGCTTACTGGCGTGGGGCC 437

701 SerGluleileuLeuproLeuProPhe.....LysGlyAspTrpIl 714
|||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| :
436 GTGCACCTATPACACCCTCTACCTTCGCGCGGAGCGGGGATTCGG 387

714 eAspGlnValArgProValIlePheIleGlyGlyGlnValPheAspT 731
|||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| :
386 GGACCTTTTCAGAACCCACTCTCTCTCAACGCTGGGAACCTC..... 344

731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
|||||:||||| ||||| ||||| ||||| ||||| ||||| ||||| :
343 .....TGCACCTGAACCTACGGGGGGGGGCCCCC 317

```

```

748  GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      ::::|::|      ::::|::|      |::|::|::|::|
316  AGGGCCCATCTGCAGCGATTGCCTGAATGT ..... 287

764  rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
286  .....ATCCGCTGGTCTCTATGGGGCGGGCTAGTTCGGCCT 250

780  yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|
249  TGGGAATATCCGACACTGGAGCTGAATATTATCCCTATGGGGTC 200

797  LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
199  CAGATGGAGACAGATCTGCGATGGGGTTCAGTTCGGCGCGGAATCCG 150

812  lPhe 813
      |||
149  ATTC 146

seq_name: gb_est1:AW115556

seq documentation block:
LOCUS AW115556 685 bp mRNA linear EST 20-OCT-1999
DEFINITION fJ97a12.x1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
259582 3' similar to SW:YDP6_SCHPO Q10478 HYPOPHYSICAL 51.8 KD
PROTEIN C17C9.06 IN CHROMOSOME 1.; mRNA sequence.
ACCESSION AW115556
VERSION AW115556.1 GI:6081894
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 685)
Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S.,
Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T.,
Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Ritter
, E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.
Washu Zebrafish EST Project 1999
Unpublished (1999)
Contact: S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T7 ET from Amersham
High quality sequence stop: 507.
Location/Qualifiers
1. 685
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="2599582"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTCGTG);
Site_2: DraIII (CAGCAGATG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTCGCTCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TCGTGGCCTACTGCG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTCG, 3' site
CAGCAGTCG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano

```

(University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA. "

BASE COUNT	192 a	179 c	145 g	169 t
ORIGIN	primer: concrdgcagcggacgac.			
alignment_scores:				
Quality:	134.50	Length:	168	
Ratio:	1.494	Gaps:	6	
Percent Similarity:	53.571	Percent Identity:	27.381	

```
alignment_block:
US-09-701-711-2 x AW115556/rev ..
Align seg 1/1 to reverse of: AW115556 from: 1 to: 685
```

```

651  GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnse 667
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
655  GACAGTTTTATTGGGTGTCCTACCAAGTGTGAGGGCTTTAGCATGTA 673
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
667  rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgargGlyGlnG 684
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
605  CAGCATCGGCCCTCAGAGTGA.. 584
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
684  lnThrThrLeuGlyGluValValcglyGlyAsnAlaLeuAlaThrPheGly 700
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
583  .....GGAGATTACCTCGGCGAGAGGCTATTGGGCTGGGGGA 545
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
701  SerGluLeuIleLeuProLeuProPhe.....LysGlyAspTrpI 714
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
544  GTTCACTGTACAGGCTCTTCCATTACGCCCTGGCGAGGTGGATTGG 495
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
714  eAspGlnValArgProValIlePheIIeGluGlyGlycInValPheAspT 731
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
494  AGATCTTTTCAGAAACACATTTCTTCCCTCAATGCCGGAACCTG..... 452
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
731  hrThrGlyMetaspLysGlnThrIleaspLeuThrGlnPheLysaspPro 747
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
451  .....TGTAACCTCAACTACGGGAAGGGCGG 425
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
 :|||:
 :|||:
 424 CGGGTCACCTCAGCAAACTGGCGGAGTGC..... 395
 :
 764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
 :|||:
 :|||:
 394ATCCGTGGTCGTAATGAGCAGGATTTCCTTCGCGTC 358
 :
 780 yThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
 :|||:
 :|||:
 357 TGGGGAAACATCGCTCGACTGGAGCTGAACACTGCATTCCTATGGGCGTC 308
 :
 797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
 :|||:
 :|||:
 307 CAGAGTGGAGACAGGATATCTGACGGAGTGCAGTTTGGGGCCGGNAATCCG 258
 :
 812 lPhe 813
 :|||
 257 CTTC 254

seq_name: qb_est1:AA474343

seq documentation block:

LOCUS	AA474343	260 bp	mrna	linear	EST 18-JUN-1997
DEFINITION	v55c03.r1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA clone IMAGE:804484 5' similar to SW:D152_HAEN P44935 PROTECTIVE SURFACE ANTIGEN D15 PRECURSOR ;, mRNA sequence.				

ACCESSION
AA474343
SOURCE ANTIEN DIO PRECURSOR,, MINNA SEQUENCE.

VERSION AA474343.1 GI:2202570

KEYWORDS EST.
SOURCE house mouse.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 260)
AUTHORS	Marra, M., Hillier, L., Allien, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
TITLE	The WashU-HMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:480828

```

FEATURES
source
Location/Qualifiers
1. .260
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone_image="804484"
/clone_lib="Knowles Solter mouse blastocyst B1"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DHI08"
/note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(gt):
5'-CGGTGCACCTGCACCGTGTGTGTGT-3', cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
75 a 58 c 68 g 59 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 134.00      Length: 86
             Ratio: 2.393      Gaps: 0
  Percent Similarity: 65.116      Percent Identity: 33.721

alignment_block:
  US-09-701-711-2 x AA474343 ..

```

```

Align seg 1/1 to: AA474343 from: 1 to: 260

289 GluLeuGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAl 305
      |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
2  GAAATTGAGCAGCTGACTAAGATCGAGCGGGTGAGCTGTATAACGGCAC 51
      |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
305 aMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspG 322
      : ||| : ||| : |||:|||| ||| : |||:|||| ||| : |||:|||| |||
52 CAAAGTACCAGAGTGGAGATGACATCAAAAGCTTCTCGTGCCTATG 101
      |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
322 lYtTYrTYrTYrAlaGlnIleArgProValThrArgIleAsnAspGluSer 338
      |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
102 GTATATGCCTATCCGGCGGTACAGTCGATCGCCGGAATTAACGATGCCAC 151
      |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
339 ArgThrValAspValGluTYrTYrIleAspProValHisProValTYrVa 355
      : |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
152 AAAACCGTTAAATTACGTCGTGAACGTTGATCGGGGTAAACCGTTTCTAGCT 201
      |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
355 lArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuA 372
      |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
202 GCGTAAGATCCGTTTGAAGGTAAAGATACCTCGAAAGATGCCGTCCTGC 251
      |||:|||| ||| |||:|||| ||| |||:|||| ||| |||:|||| |||
372 rgArgGlu 374

```

|||||
252 GTCCGAA 259

seq_name: gb_est2:BM119288

seq_documentation_block:
LOCUS BM119288 697 bp mRNA linear EST 01-FEB-2002
DEFINITION L0923D10-3 NTA Mouse Newborn Kidney cDNA Library (Long) Mus
musculus cDNA clone L0923D10 3', mRNA sequence.

ACCESSION BM119288
VERSION BM119288.1 GI:17087314
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 697)

REFERENCE Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Pantano, S., Lim, M.K.
AUTHORS and Ko, M.S.H.

TITLE Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library
JOURNAL Unpublished (2001)
COMMENT Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@igsun.grc.nia.nih.gov

Plate: L0923 row: D column: 10
Seq primer: -21M13 Forward
High quality sequence stop: 697
POLYA-res.

FEATURES
source Location/Qualifiers
1..697
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="nia:EST:L0923D10-3"
/db_xref="taxon:10090"
/clone="L0923D10"
/clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long)"
/tissue_type="Newborn Kidney"
/dev_stage="Newborn"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pSPORT1 (Invitrogen);
Site_1: SalI; Site_2: NotI; Mouse cDNA project by the
Laboratory of Genetics, National Institute on Aging (NIA),
Intramural Research program, NIH
(http://igsun.grc.nia.nih.gov/cDNA). This is a
long-transcript enriched cDNA library (Ref. Genome Res.
11:1553-1558 (2001). [PMID:11544199]). In brief
, double-stranded cDNAs were synthesized with an Oligo(dT)
primer [Invitrogen: 5'-
pGACTAGTCTAGATCGAGCGCGCCGCTTTT-3'] from 26
microgram of total RNA, treated with T4 DNA polymerase,
and purified by ethanol-precipitation. The cDNAs were
ligated to lone-linker LL-SalI, purified by
phenol/chloroform, and separated from free linkers by
Centricon 100. Then, the cDNAs were amplified by
long-range high fidelity PCR using Ex Taq polymerase
(Takara) with a primer Sal4-S. The products were purified
by phenol/chloroform and Centricon 100. The cDNAs were
digested with SalI and NotI enzymes, and cloned into
SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
coli host was transformed with ligation mixture by the
standard chemical method. The average insert size is about
3.0 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT 159 a 207 c 185 g 146 t

ORIGIN
alignment_scores:
Quality: 134.00 Length: 195
Ratio: 1.354 Gaps: 8
Percent Similarity: 50.769 Percent Identity: 26.154

alignment_block:
US-09-701-711-2 x BM119288/rev ..
Align seg 1/1 to reverse of: BM119288 from: 1 to: 697

639 AlaLysLeuGlyTyrGly.....AsnAsnLe 647
|||||
621 GTCCGCTTGGACTCGGTATTCTCCACGCTCTCTGGGTGGAATGCT 572
647 uPro.....PheTyrGluAsnPheTyrAlaGly 657
|||||
571 GCCCATCGGTGACAGCCATTCAGCATTCGTGACAGGTTTACCTGGAG 522
657 lYtYrGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSer 673
|||||
521 GCCCCACGAGTGTCCGAGGATTAGCATGCACAGCATTCGACCCAGAGT 472
674 GlnAlaTyrLeuThrAlaArgGlyGlnGlnThrThrLeuGlyGluVa 690
|||
471 GAA.....GGAGATTA 461
690 lValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProL 707
|||||
460 CTGGCGCGGAGGCTACTGGCTGGGGCTGCACCTCTACACCCAC 411
707 euProPheLys.....GlyAspTyrPheAspGlnValArgProVal 720
|||||
410 TGCCTTCGGCCAGCCAGGTGGCTTCGGAGAGCTTTTCAGAACTCAC 361
721 lIePheIleGluGlyGlnValPheAspThrThrGlyMetAspLysGl 737
|||||
360 TTTTCTCAATCGGGCAACCTG..... 337
737 nThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnA 754
|||||
336 ...TGCAACCTCAACTATGCTGGGGCCCCAAAGCCCATATCCGGAAGC 291
754 snAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLysGlnLeuArg 770
|||||
290 TAGCTGAGTGC.....ATCCGC 274
771 TyrSerAlaGlyValGlyAlaThrTrp...TyrThrProIleGlyProLe 786
|||||
273 TGTCTCTATGAGCAGCGCTCTCTCCGACTTGCCACATCGCTCGGT 224
786 uSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAspGln...T 802
|||||
223 GGAGCTGAACTACTGCAATTCCTATGGCGCTGCAGGGGGCGGACAGATT 174
802 hrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
173 GTGATGGTGTCCAGTTTGGAGCTGGGATTCGGTTC 139

seq_name: gb_est2:BG303393

seq_documentation_block:

LOCUS BG303393 675 bp mRNA linear EST 23-FEB-2001
DEFINITION f155a04.x1 Sugano Kawakami zebrafish DRB Danio rerio cDNA clone
3817615 3' similar to TR:Q9Y512 Q9Y512 DJ796117.2 ; mRNA sequence.

ACCESSION BG303393
VERSION BG303393.1 GI:13100920
KEYWORDS EST.
SOURCE zebrafish.

ORGANISM Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 675)

AUTHORS Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S.,
Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T.,
Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Ritter,
E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE Washu Zebrafish EST Project 1999

pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close lab (Choi) at the University of California, Riverside. Also Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing) plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)"

BASE COUNT 246 a 154 c 197 g 207 t
ORIGIN

alignment_scores:

Quality: 133.50 Length: 271
Ratio: 1.027 Gaps: 8
Percent Similarity: 47.970 Percent Identity: 19.557

alignment_block:

US-09-701-711-2 x BG344329 ..

Align seg 1/1 to: BG344329 from: 1 to: 804

```

281 GlyAsnLeuThrThrGlnAlaGluLeuGluAlaLeuLeuLysPheY 297
||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 GGCATGTTTCATATGCTGAGATCCITTCGGAGGAATTCCTAGGCTGCA 54

297 sAlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnI 314
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
55 AGTTCCTGAAGCC.....GAGGTTTAATAATA 80

314 leSerThrLysPheGlyAspGlyThrTyrrTyrrAlaGlnIleArgPro 330
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
81 TCAATATCCCTTCTCGAT..... 100

331 ValThrArgIleAsnAspGluSerArgThrValAspValGluTyrrIl 347
100 ..... 100

347 eAspProValHisProValTyrrValArgArgIleAsnPheThrGlyAsnP 364
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
101 .....AGAGAATCTGGTGAACCAACTACTG 125

364 heLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGly 380
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
126 GAAAAAGAAACCAACCACTACTCCAGCAGCTGACCACCAAGAGGCT 175

381 AlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetAr 397
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
176 CAGGCATACATAGACACAGCAAGTGAAAGAGAGATGTTGAAACGATAC 225

397 gThrGlyPhePheLysHisValThrValAspThrArgProValProAsnS 414
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
226 TATCGGGAATATGAGGATGTCACATAATATCCGACCCCTGTGGGAGATT 275

414 erProAspGlnValAspValAsnPheValValGluGlnProSerGly 430
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
276 CT...AATAAGTGGATCTTGTTCATGAATCTTGTGCAACCAAAATCTGCT 322

431 SerSerThrIleAlaAlaGlyTyrrSerGlnSer..... 441
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
323 GGTTCCTCTGCTGCTGGTGGCATTTCCAGTGGGATAACAAATGGGCTCT 372

442 .GlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetClyt 458
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

373 TTCTGGATTAAATGGCAGCTTTGCATATCTCAGCGGAATGTTTTGGGA 422
458 hrClyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVal 474
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
423 GGAACAAGAGTTGAATCTCTCACTAGAGGGGCAAAATAGATTCTATA 472
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
475 TyrSerLeuGlyMetThrAsnProTyrrPheThrValAsnGlyValSerGI 491
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
473 TTTCGTTTAAACTACACTGACCTCTGG.....ATTGATGGT..... 508
491 nSerLeuSerGlyTyrrTyrrArgLysThrLysTyrrAspAsnLysAsnIleS 508
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
509 .....GACAAATAAGAGGACCT 524
508 eRAsnTyrrValLeuAspSerTyrrGlySerLeuSerTyrrGlyTyrrPro 524
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
525 CCAGAACTGTCTATG..... 538
525 IleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspAsnThrLy 541
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
539 ...GTTCAAAACTCTAGG.....ACCCCTGGAACACT 567

541 sLeuHisGlyGly 545
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
568 TATCCATGTTGGG 580

seq_name: gb_est2:BI278122

seq_documentation_block:
LOCUS BI278122 573 bp mRNA linear EST 19-JUL-2001
DEFINITION UI-R-CZ0-byd-f-02-0-UI.s1 UI-R-CZ0 Rattus norvegicus CDNA clone
UI-R-CZ0-byd-f-02-0-UI 3', mRNA sequence.
ACCESSION BI278122
VERSION BI278122.1 GI:14924662
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 573)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized penis library cDNA Library Preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.
FEATURES
source
1..573
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-CZ0-byd-f-02-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pMT73D-Pac (Pharmacia) with a modified

```


polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C20 library is a non-normalized library constructed from rat penis tissue. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_LIB=UI-R-C20
TAG_TISSUE=penis
TAG_SEQ=TTGCGGAACA"

BASE COUNT 126 a 164 c 153 g 130 t
ORIGIN

alignment_scores:

Quality: 132.50 Length: 168
Ratio: 1.472 Gaps: 6
Percent Similarity: 53.571 Percent Identity: 26.786

alignment_block:

US-09-701-711-2 x BI278122/rev ..

Align seg 1/1 to reverse of: BI278122 from: 1 to: 573

```

651 GluAsnPheTyrAlaGlyClyTyrGlySerValArgGlyTyrAspGluSe 667
      ::::::::::::::::::::|
558 GACAGGTTTACCTGGAGGCCACGAGCGTCCGAGGATTAGCATGCA 509
      ::::::::::::::::::::|
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      ::::::::::::::::::::|
508 CAGCATCGGACCCAGAGCGAA..... 487
      ::::::::::::::::::::|
684 InThrLeuGlyValGlyValGlyAsnAlaLeuAlaThrPheGly 700
      ::::::::::::::::::::|
486 .....GGAGATTACCTGGCGGGAGGCATACCTGGCGCGGGGGC 448
      ::::::::::::::::::::|
701 SerGluLeuIleLeuProLeuProPheLys.....GlyAspTrpI 714
      ::::::::::::::::::::|
447 CTGCACCTCTATACCCACCTGCTTTCGGCCAGGCAGGTTGGCTTCGG 398
      ::::::::::::::::::::|
714 eAspGlnValArgProValIlePheIleGlyGlyGlnValPheAspT 731
      ::::::::::::::::::::|
397 AGAGCTTTTCAGACTCACTTTTCTCAATGCTGGCAACCTG..... 355
      ::::::::::::::::::::|
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
      ::::::::::::::::::::|
354 .....TGCACCTCACTAGCTGGTGGAGGCCCC 328
      ::::::::::::::::::::|
748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      ::::::::::::::::::::|
327 AGAGCCACATCCCGAAGCTGGCTCAATGC..... 298
      ::::::::::::::::::::|
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
      ::::::::::::::::::::|
297 .....ATCCGCTGGTCTATGGTGCAGGCATCGCTCCCGCC 261
      ::::::::::::::::::::|
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
      ::::::::::::::::::::|
260 TTGGCAACATTCCTCGGTGGAGTGAACACTACTGCATTCTTATGGGTGTG 211
      ::::::::::::::::::::|
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
      ::::::::::::::::::::|
210 CAGCGAGCGACAGGATTGTGTGGCTCCAGTTTGGAGCTGGGATTCCG 161
      ::::::::::::::::::::|
812 lPhe 813
      ::::::::::::::::::::|
160 GTTC 157

```

seq_name: gb_est2:BJ054028

seq_documentation_block:

LOCUS BJ054028 721 bp mRNA linear EST 11-DEC-2001
DEFINITION BJ054028 NTBB Mochii normalized Xenopus neurula library

laevis cDNA clone X1049a15 3', mRNA sequence.

BJ054028

VERSION BJ054028.1 GI:17499034

KEYWORDS EST

SOURCE African clawed frog.

ORGANISM Xenopus laevis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 721)

AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara

,Y.

TITLE Expressed genes in X. laevis embryo

JOURNAL Unpublished (2001)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES

Location/Qualifiers

1..721

/organism="Xenopus laevis"

/db_xref="taxon:8355"

/clone="X1049a15"

/clone_lib="NTBB Mochii normalized Xenopus neurula

library"

/tissue_type="whole embryo"

/dev_stage="stage 15"

BASE COUNT 162 a 218 c 183 g 157 t 1 others

ORIGIN

alignment_scores:

Quality: 132.00 Length: 214
Ratio: 1.211 Gaps: 8
Percent Similarity: 50.935 Percent Identity: 26.168

alignment_block:

US-09-701-711-2 x BJ054028/rev ..

Align seg 1/1 to reverse of: BJ054028 from: 1 to: 721

```

612 GlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgPr 628
      ::::::::::::::::::::|
687 GGGGACGTCAGTTTCTTAAGGAAGACTTTGAGTTGCAGTTGAACAGGCA 638
      ::::::::::::::::::::|
628 oPheIleLysLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyA 645
      ::::::::::::::::::::|
637 GCTTACCTGGGACTCTGCTGTCG.....ACATCTCTGTGGGGAGAA 594
      ::::::::::::::::::::|
645 snAsnLeuPro.....PheTyrGluAsnPheTyr 654
      ::::::::::::::::::::|
593 TGTGTGTCGCGATCGGAGACAGCGCCGACGAGTATCGACACCGTTCTAC 544
      ::::::::::::::::::::|
655 AlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyPr 671
      ::::::::::::::::::::|
543 CTGGGGGGACCCACGAGCGTGGAGGATTCAGCATGTACAGCATGGCCCC 494
      ::::::::::::::::::::|
671 oArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuG 688
      ::::::::::::::::::::|
493 CCAGAGTGAA..... 483
      ::::::::::::::::::::|
688 lylGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuIle 704
      ::::::::::::::::::::|
482 GTGACTATTGGCGGGGAAGCTTACTGGGCTGGGGCGTGCACATTATAC 433
      ::::::::::::::::::::|
705 LeuProLeuProPhe.....LysGlyAspTrpIleAspGlnValAr 718
      ::::::::::::::::::::|
432 ACCCCCTCCCTTCGGCCGGAGCGCGGGGATTCGGGGACCTTTTTCAG 383
      ::::::::::::::::::::|
718 gProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyMeta 735

```

```

382 GACTCACTTCTCTGGAACGCTGGGAACCTC..... 352
735 spLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAla 751
351 .....TGCACACTTGAACTACGGGGAGGGCCCGGGGGCCCATCTT 313
752 GluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLysG1 768
312 CGACGATTGGCTGAATGT..... 295
768 nLeuArgTyrSerAlaGlyValGlyAlaThrTrp...TyrThrProIleG 784
294 .ATCCGCTGGTCCATATGCGGGGGGGCTAGTTCTGCGCTTGGGGAATATCG 246
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
245 CGAGACTGGAACGAAATATTATGATCCCTATGCGGGGTCCAGAGCGGAGAC 196
801 Gln...ThrAspThrValGlnPheGlnIleGlySerValPhe 813
195 AGGATCTCGCAGCGGGTCCAGTTCCGGCGCGGAATCCGATTC 154

```

seq_name: gb_est2:BJ088421

seq_documentation_block:
LOCUS BJ088421 481 bp mRNA linear EST 12-DEC-2001
DEFINITION BJ088421 N1BB Mochii normalized xenopus tailbud library Xenopus laevis cDNA clone XL093a01 3', mRNA sequence.

ACCESSION BJ088421
VERSION BJ088421.1 GI:17585682

KEYWORDS EST.
SOURCE African clawed frog.

ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 481)

REFERENCE
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara Y.

TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
Location/Qualifiers

source
1..481
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XL093a01"
/library="N1BB Mochii normalized xenopus tailbud library"
/tissue_type="whole embryo"
/dev_stage="stage 25"
98 a 161 c 132 g 90 t

alignment_scores:
Quality: 131.50 Length: 168
Ratio: 1.478 Gaps: 6
Percent Similarity: 52.976 Percent Identity: 27.381

alignment_block:

US-09-701-711-2 x BJ088421/rev ..

Align seg 1/1 to reverse of: BJ088421 from: 1 to: 481

651 GluAsnPheThrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667

```

438 GACGGTCTTACCTGGGGGACCCACGAGCTGCGAGGATTCAGCATGTA 389
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaAargArgGlyGlnG 684
388 CAGCATTGGCCCCCAGAGTGAA..... 367
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
366 .....GGTGACTATTGGCGGGGAAGCTTACTGGCGTGGGGGC 328
701 SerGluLeuIleLeuProLeuProPhe.....LysGlyAspTrpI1 714
327 GTGCACATTATACACCCCTCCCTTCGGCGCGGACGGGGGATTCGG 278
714 eAspGlnValargProValIlePheIleGluGlyGlyGlnValPheAspT 731
277 GGACCTTTTTCAGGACTCATTCTCTCAGCGCTGGGAACCTC..... 235
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
234 .....TGCACTTGAACCTACGGGGGAGGGCCCC 208
748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
207 GGGGCCCATCTTCGACGATTGGCTGAATGT..... 178
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
177 .....ATCCGCTGGTCTATGGGGCGGGCTAGTTCTGCGCT 141
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
140 TGGGGAATATCGGAGACTGGAAGTGAATATTGATCCCTATGGGGGTC 91
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
90 CAGAGCGGAGACAGGATCTCGAGGGGTCCAGTTCCGGCGCGGAATCCG 41
812 lPhe 813
40 ATTC 37

```

seq_name: gb_est2:BM233570

seq_documentation_block:

LOCUS BM233570 580 bp mRNA linear EST 31-JAN-2002
DEFINITION K0338G02-3 N1A Mouse Osteoblast cDNA Library (Long) Mus musculus CDNA clone K0338G02 3', mRNA sequence.

ACCESSION BM233570
VERSION BM233570.1 GI:17868840

KEYWORDS EST.
SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE

AUTHORS Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Carter,M.G., Umezawa,A. and Ko,M.S.H.

TITLE Systematic Analyses of N1A Mouse Osteoblast cDNA Library (Long)

JOURNAL Unpublished (2001)

COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdnae@igsun.grc.nia.nih.gov
Plate: K0338 row: G column: 02
Seq primer: -21M13 Forward
High quality sequence stop: 580
POLYA=Yes.

FEATURES
source Location/Qualifiers
1..580 /organism="Mus musculus"

```

/strain="C3H/He mice"
/db_xref="niaEST:K0338602-3"
/db_xref="taxon:10090"
/clone_lib="NIA Mouse Osteoblast cDNA Library (Long)"
/clone_lib="K0338602"
/tissue_type="Osteoblast"
/cell_line="KUSA-A1 cells"
/lab_host="DH10B"
/notes="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001). [PMID: 1154199]). Total RNAs were
obtained from Dr. Akihiro Umezawa (Keio University School
of Medicine, Japan). Double-stranded cDNAs were
synthesized with an Oligo(dT) primer [Invitrogen:
5'-pGACATGTCATGATCGGCGGCCCTTTT-3'] from
2.1 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lona-linker LI-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPORT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 3.0 kb. The library was constructed
by Yulan Piao (NIA)."
BASE COUNT      131 a 174 c 156 g      119 t
ORIGIN

```

```

alignment_scores:
  Quality: 131.50      Length: 168
  Ratio: 1.461        Gaps: 6
  Percent Similarity: 53.571      Percent Identity: 26.786

alignment_block:
US-09-701-711-2 x BM233570/rev ..

Align seg 1/1 to reverse of: BM233570 from: 1 to: 580

651 GluAsnPheTyrAlaGlyGlyTyrSerValArgGlyTyrAspGlnSe 667
   :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
490 CAGCATTTGGACCCAGAGTGAA..... 469
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
468 .....GGAGATTACCTGGGCGCGGAGGCTACTGGCTGGGGGC 430
701 SerGluLeuLeuLeuProLeuProPheLys.....GlyAspTrpI 714
   ::|||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
429 CTGCACCTCTACACCCACCTGCTCCGCGCAGGCGAGGCTGCTCGG 380
714 eAspGlnValArgProValPhePheLeuGlyGlyGlnValPheAspT 731
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
379 AGAGCTTTTCAGAACTCACTTTTTCCTCAATGCGGCAACCTG..... 337
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
336 .....TGCAACCTCACTATGCTGAGGCGCC 310
748 GlnAlaThrAlaGluGlnAsnAlaLysAlaAsnArgProLeuLeuTh 764
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
309 AAAGCCCATTCGGAAGTAGCTGAGTGC..... 280

```

```

764 rclnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...T 780
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
279 .....ATCCGCTGCTCTATGGAGCAGCGTGGCTCCGAC 243
780 yrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
242 TTGCAGCATCGCTGGCTGAGCTGAACTACTGCATTCCTATGGCGTG 193
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySerVa 812
   ::|||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
192 CAGGGGGCGCAGCAGGATTGTGATGCTGCCAGTTTGGAGCTGGGATTCG 143
812 lphe 813
   |||
142 GTTC 139

seq_name: gb_est1:BB201163

seq_documentation_block:
LOCUS      BB201163                641 bp      mRNA      linear      EST 19-OCT-2001
DEFINITION BB201163 RIKEN full-length enriched, 0 day neonate thymus Mus
            musculus cDNA clone A430034M11 3', mRNA sequence.
ACCESSION  BB201163
VERSION    BB201163.2 GI:16271538
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 641)
AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
            Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
            M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
            D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
            RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
            Unpublished (2001)
TITLE      On Jun 30, 2000 this sequence version replaced gi:8866116.
JOURNAL    Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsr.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,K., Matsuura
            S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multipipillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            Y., and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.

```

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues

FEATURES

Location/Qualifiers
 1. .641
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="A430034M11"
 /clone_lib="RIKEN full-length enriched, 0 day neonate thymus"
 /tissue_type="thymus"
 /dev_stage="0 day neonate"
 /lab_host="DH10B"
 /note="Site:1: SalI; Site:2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATAATATCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 129 a 170 c 190 g 152 t
 ORIGIN

alignment_scores:

Quality: 131.00 Length: 199
 Ratio: 1.272 Gaps: 8
 Percent Similarity: 51.759 Percent Identity: 26.131

alignment_block:

US-09-701-711-2 x BB201163 ..
 Align seg 1/1 to: BB201163 from: 1 to: 641

```

627 ArgProPheIleLysSerValLeuArgGlyTyrAlaLysLeuGlyTy 643
      ::::::::::::::::::::
15  AAGCCGTGCGCTTGGACTCGGTATTTCC.....ACGTCTCTCTGGGG 58
      ::::::::::::::::::::
643 rGlyAsnAsnLeuPro.....PheTyrGluAsnP 653
      ::::::::::::::::::::
59  TGGATGCTGTGGCCATCGGTGACAGCCATCCAGCATTTGTGACAGT 108
      ::::::::::::::::::::
653 heTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSerLeu 669
      ::::::::::::::::::::
109  TTTACTGGAGGCCCGGAGGTGCGGAGGATTAGCATGCACAGCAT 158
      ::::::::::::::::::::
670 GlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnThrTh 686
      ::::::::::::::::::::
159  GGACCCGAGAGTGAA..... 173
      ::::::::::::::::::::
686 rLeuGlyGluValValGlyValGlyAsnAlaLeuAlaThrPheGlySerGluL 703
      ::::::::::::::::::::
174  ....GGGATTACCTGGGGCGGAGCCCTACTGGGCTGGGGCGCTGCACC 219
      ::::::::::::::::::::
703 euIleLeuProLeuProPheLys.....GlyAspTrpIleAspGln 716
      ::::::::::::::::::::
220  TTTACACCCCACTGCCCTTCGGCCAGGCCAGGCGTGGCTTCGGAGAGCT 269
      ::::::::::::::::::::
717 ValArgProValIlePheIleGlyGlyGlnValPheAspThrThrGl 733
      ::::::::::::::::::::
270  TTCAGACTCACTCTTCTCAATCGGGCAACCTCG..... 305

```

```

733 yMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaT 750
      ::::::::::::::::::::
306  ....TGCACCTCACTATGTTGGTGGGGCCCCCAAGCCC 339
      ::::::::::::::::::::
750 hrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAsp 766
      ::::::::::::::::::::
340  ATATCCGGAAGCTAGCTGAGTGC..... 362
      ::::::::::::::::::::
767 LysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...TyrThrPr 782
      ::::::::::::::::::::
363  ....ATCCGCTGTGCTATGAGCAGCGCTCTCCCGACTTGGCAA 406
      ::::::::::::::::::::
782 oIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysGlnA 799
      ::::::::::::::::::::
407  CATCCTCGGTGGAGCTGAAGTACTGCTATGCTATGGCGTGCAGGGG 456
      ::::::::::::::::::::
799 snAspGln...ThrAspThrValGlnPheGlnIleGlySerValPhe 813
      ::::::::::::::::::::
457  GCGACAGGATTTGTGATGTGTCCAGTTTGGAGCTGGGATTCGGTTC 503

```

seq_name: gb_est2:BI539548

seq_documentation_block:

LOCUS BI539548 548 bp mRNA linear EST 30-AUG-2001
 DEFINITION 452266 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BI539548
 VERSION BI539548.1 GI:15380658
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
 Bovidae; Bovinae; Bos.
 Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
 ,G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G.,
 Perlea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
 Keele,J.W.

REFERENCE

AUTHORS

Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCCAGTCACGACG

Plate: 114 row: B column: 1

Seq primer: ATTTAGTGCACATATAG.

FEATURES

source

Location/Qualifiers
 1. .548
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 1BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site:1: XbaI; Site:2: XhoI;
 Library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."
 BASE COUNT 103 a 181 c 168 g 96 t
 ORIGIN

alignment_scores:

Harvard University, Howard Hughes Medical Institute
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@tothp.harvard.edu
 Pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,
 2000) library was constructed by Catherine Lee DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Marie Searce
 (mssearce@mail.med.upenn.edu)
 Seq primer: -40UP from Gibco
 High quality sequence stop: 423.

FEATURES	SOURCE
1. The first two columns are labeled "FEATURES" and "SOURCE".	
2. The third column is labeled "FEATURES".	
3. The fourth column is labeled "SOURCE".	
4. The fifth column is labeled "FEATURES".	
5. The sixth column is labeled "SOURCE".	
6. The seventh column is labeled "FEATURES".	
7. The eighth column is labeled "SOURCE".	
8. The ninth column is labeled "FEATURES".	
9. The tenth column is labeled "SOURCE".	
10. The eleventh column is labeled "FEATURES".	
11. The twelfth column is labeled "SOURCE".	
12. The thirteenth column is labeled "FEATURES".	
13. The fourteenth column is labeled "SOURCE".	
14. The fifteenth column is labeled "FEATURES".	
15. The sixteenth column is labeled "SOURCE".	
16. The seventeenth column is labeled "FEATURES".	
17. The eighteenth column is labeled "SOURCE".	
18. The nineteenth column is labeled "FEATURES".	
19. The twentieth column is labeled "SOURCE".	
20. The twenty-first column is labeled "FEATURES".	
21. The twenty-second column is labeled "SOURCE".	
22. The twenty-third column is labeled "FEATURES".	
23. The twenty-fourth column is labeled "SOURCE".	
24. The twenty-fifth column is labeled "FEATURES".	
25. The twenty-sixth column is labeled "SOURCE".	
26. The twenty-seventh column is labeled "FEATURES".	
27. The twenty-eighth column is labeled "SOURCE".	
28. The twenty-ninth column is labeled "FEATURES".	
29. The thirtieth column is labeled "SOURCE".	
30. The thirty-first column is labeled "FEATURES".	
31. The thirty-second column is labeled "SOURCE".	
32. The thirty-third column is labeled "FEATURES".	
33. The thirty-fourth column is labeled "SOURCE".	
34. The thirty-fifth column is labeled "FEATURES".	
35. The thirty-sixth column is labeled "SOURCE".	
36. The thirty-seventh column is labeled "FEATURES".	
37. The thirty-eighth column is labeled "SOURCE".	
38. The thirty-ninth column is labeled "FEATURES".	
39. The fortieth column is labeled "SOURCE".	
40. The forty-first column is labeled "FEATURES".	
41. The forty-second column is labeled "SOURCE".	
42. The forty-third column is labeled "FEATURES".	
43. The forty-fourth column is labeled "SOURCE".	
44. The forty-fifth column is labeled "FEATURES".	
45. The forty-sixth column is labeled "SOURCE".	
46. The forty-seventh column is labeled "FEATURES".	
47. The forty-eighth column is labeled "SOURCE".	
48. The forty-ninth column is labeled "FEATURES".	
49. The fiftieth column is labeled "SOURCE".	
50. The fifty-first column is labeled "FEATURES".	
51. The fifty-second column is labeled "SOURCE".	
52. The fifty-third column is labeled "FEATURES".	
53. The fifty-fourth column is labeled "SOURCE".	
54. The fifty-fifth column is labeled "FEATURES".	
55. The fifty-sixth column is labeled "SOURCE".	
56. The fifty-seventh column is labeled "FEATURES".	
57. The fifty-eighth column is labeled "SOURCE".	
58. The fifty-ninth column is labeled "FEATURES".	
59. The sixtieth column is labeled "SOURCE".	
60. The sixty-first column is labeled "FEATURES".	
61. The sixty-second column is labeled "SOURCE".	
62. The sixty-third column is labeled "FEATURES".	
63. The sixty-fourth column is labeled "SOURCE".	
64. The sixty-fifth column is labeled "FEATURES".	
65. The sixty-sixth column is labeled "SOURCE".	
66. The sixty-seventh column is labeled "FEATURES".	
67. The sixty-eighth column is labeled "SOURCE".	
68. The sixty-ninth column is labeled "FEATURES".	
69. The seventieth column is labeled "SOURCE".	
70. The seventy-first column is labeled "FEATURES".	
71. The seventy-second column is labeled "SOURCE".	
72. The seventy-third column is labeled "FEATURES".	
73. The seventy-fourth column is labeled "SOURCE".	
74. The seventy-fifth column is labeled "FEATURES".	
75. The seventy-sixth column is labeled "SOURCE".	
76. The seventy-seventh column is labeled "FEATURES".	
77. The seventy-eighth column is labeled "SOURCE".	
78. The seventy-ninth column is labeled "FEATURES".	
79. The eightieth column is labeled "SOURCE".	
80. The eighty-first column is labeled "FEATURES".	
81. The eighty-second column is labeled "SOURCE".	
82. The eighty-third column is labeled "FEATURES".	
83. The eighty-fourth column is labeled "SOURCE".	
84. The eighty-fifth column is labeled "FEATURES".	
85. The eighty-sixth column is labeled "SOURCE".	
86. The eighty-seventh column is labeled "FEATURES".	
87. The eighty-eighth column is labeled "SOURCE".	
88. The eighty-ninth column is labeled "FEATURES".	
89. The ninetieth column is labeled "SOURCE".	
90. The ninety-first column is labeled "FEATURES".	
91. The ninety-second column is labeled "SOURCE".	
92. The ninety-third column is labeled "FEATURES".	
93. The ninety-fourth column is labeled "SOURCE".	
94. The ninety-fifth column is labeled "FEATURES".	
95. The ninety-sixth column is labeled "SOURCE".	
96. The ninety-seventh column is labeled "FEATURES".	
97. The ninety-eighth column is labeled "SOURCE".	
98. The ninety-ninth column is labeled "FEATURES".	
99. The hundredth column is labeled "SOURCE".	

```

FEATURES
Location/Qualifiers
1. 570
/organism="Mus musculus"
/strain="129/Sv x CD1"
/db_xref="taxon:10090"
/clone_lib="kaestner ngn3 wt"
/dev_stage="p.c. 14.5"
/lab_host="E. coli-DH12S (GIBCO)"
/note="Organ: pancreas; Vector: pNot1; Site.2: Sal I; The library was obtained from Gerard Gradwohl 2000". The cDNA's were prepared w
Not1 site, and SalI linkers were inserts were cut with NotI before Not1-SalI sites in the vectors. T libraries, ngn3 wt and ngn3 -/-, pSPORT1, T7 promoter is 5'."
127 a 165 c 149 q 129 t
BASE COUNT

```

BASE COUNT	127 a	165 c	149 g	129 t
ORIGIN				
alignment_scores:				
Quality:	130.50			Length: 168
Ratio:	1.450			Gaps: 6
Percent Similarity:	53.571			Percent Identity: 26.786

Align seq 1/1 to reverse of: BI714058 from: 1 to: 570

```

651  GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
555  GACAGGTTTTACTTGGGAGGCCCCACGAGTGTCCGAGGATTTAGCATGCA 506
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
667  rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
505  CAGCATTTGGACCCCGAGAGTGAA..... 484
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
684  lnThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
      | | | | | : : | | | | | | | | | | | | | | | | | | | | | |
483  .....GGAGATTACCTGGCGGCGAGGCTTACTGGGCTGGGGCG 445
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
701  SerGluLeuLeuProLeuProPheLys.....GlyAspTrpIl 714
      : : | | | | | | | | | | | | | | | | | | | | | | | | | |
444  CTGCACCTCTACACCCCACTGCCTTCGCGCAGCGCCAGGTGGCTCGG 395
      : : : : : | | | | | | | | | | | | | | | | | | | | | |
714  eAspGlnValArgProValIlePheIleGluGlyGlyClnValPheAspT 731
      : : : | | | | | | | | | | | | | | | | | | | | | | | | | |
394  AGAGCTTTTCAGAACTCACATTTTTCCTCAATGCGGGCAACCTG..... 352
      : : : | | | | | | | | | | | | | | | | | | | | | | | | | |
731  hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
      : : : | | | | | | | | | | | | | | | | | | | | | | | | | |
351  .....TGCAACCTCAACTATGGTCAGGGCCCC 325
      : : : | | | | | | | | | | | | | | | | | | | | | | | | | |
748  GlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      : : | | | | | : : : : : | | | | | | | | | | | | | | | | |
324  AAAGCCCATATCCGAGAGCTAGCTGAGTGC..... 295

```

Quality:	130.50	Length:	166
Ratio:	1.483	Gaps:	6
Percent Similarity:	53.012	Percent Identity:	27.108

alignment_block:

US-09-701-711-2 x BI539548

Align seq 1/1 to: BI539548 from: 1 to: 548

653	pheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSeriLe	669
2	TtCTACCTTGCGGCACCACCGAGGTGGGGCTTCAGCATCAGCGGT	51
669	wGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnGlnThrT	686
:	:	:
52	CGGGCCACAGAGCGAA	67
686	hrLeuGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGlu	702
68GGTAGTACTGGCGGGGAAGCGCTACTGGCGCGTGGGTGCAC	112
703	LeuIleLeuProLeuProPheLys.....GlyAspTrpIleAspGI	716
113	CTGTACACCCCGTTTACCGTTCGGCGCGGCCAGGCGGCTTTGGCGAGCT	162
716	nValArgproValIlePheIleGluGlyGlnValPheAspThrThrG	733
163	CTTCAGGACACACTTCTCCTCAACGGGGGAACCTC.....	199
733	lyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAla	749
200TGCAACCTCAACTACGGGGAGGCCCAAGCC	232
750	ThrAlaGluGlnAsnAlaLysAlaAAsnArgProLeuLeuThrGlnAs	766
233	CACATCCGGGAAGCTGGCGGATGC.....	256
766	pLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp...TyrThrP	782
:	:	:
257ATCCGCTGGTTCACGGGGGGGGCATCGTCTCAGCGCTGGGCA	299
782	roileGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysGln	798
300	ACATCGCGCGCTGGAGCTCAACTACTCGCTCCCATGGCGGTGAGCGGT	349
799	AnaAspGln...ThrAspThrValGlnPheGlnIleGlySerValphe	813
:	:	:
350	GGGGACAGGACTCTGTACGGCGCTCCAGTTCGTTCGAGCTGGGATCAGGTC	397

seq name: qb est2:BI714058

seq_documentation_block:					
LOCUS	BI714058	570 bp	mRNA	linear	EST 19-SEP-2001
DEFINITION	ie31f12.x1 Kaestner ngn3 wt Mus musculus cDNA 3', similar to				
	TR:Q9Y512 Q9Y512 DJ796117.2 ;, mRNA sequence.				
ACCESSION	BI714058				
VERSION	BI714058.1	GI:15689753			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 570)				
AUTHORS	Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scearcia, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y. and Bowers, Y.				
TITLE	Endocrine Pancreas Consortium				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue				
	Endocrine Pancreas Consortium				


```

/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer (5'
TCGTACCACTGTAAGTGGGAGCGCCGATGAGTTTGTGTGTGTGTGTGTGT
T 3'); double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

```

```

BASE COUNT      116 a      162 c      161 g      151 t      1 others
ORIGIN

```

```

alignment_scores:
  Quality: 130.50      Length: 168
  Ratio: 1.466        Gaps: 6
  Percent Similarity: 52.976      Percent Identity: 26.190

```

```

alignment_block:
US-09-701-711-2 x A1553725/rev ..

```

```

Align seg 1/1 to reverse of: A1553725 from: 1 to: 591

```

```

651 GUAasnPhetYrAlaGlyTYrGlySerValArgGlyTYrAspGlnse 667
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
573 GATAGGTTTAACTTGGGGACCACCAAGCATCCGCGATTACACATGCA 524
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGln 684
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
523 CAGCATGGGGCCACAGACCA..... 502
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
684 InThrThrLeuGlyGlnValAlaGlyAlaLeuAlaThrPheGly 700
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
501 .....GGAGACTACCTAGTGGAGAACGCTAGCGCGCGCGC 463
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
701 SerGlnLeuLeuProLeuProPheLys.....GlyAspTrrpII 714
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
462 CTGACCTCTACACCCCATTCCTTCGGCAGGCCAGCGGTGGCTTGG 413
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
714 eaSpGlnValArgProValIlePheIleGlyGlyGlnValPheAsp 731
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
412 AGAACTTTCGGAACACATCTTCTCAACGACGAAACCTC..... 370
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAsp 747
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
369 .....TGCAACTCAACTATGGGAGGCGCC 343
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
748 GlnAlaThrAlaGlnAlaAsnAlaLysAlaAlaAsnArgProLeuLeu 764
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
342 AAGCTCATTTGCTAAGCTGCTGAGTGC..... 313
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrp... 780
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
312 .....ATCCGCTGTGTAAGCGCGGAGATGTCTCAGGC 276
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
780 TyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsn 796
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
275 TTGGCAACATCGCTGCGTGAACCTTAATTAATTCGCTCCCATGGAG 226
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleGlySer 812
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
225 CAGACAGGGGACAGCATATGTATGCGCTCCACTTTGGAGCTGGGATAG 176
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
812 lPhe 813
      |||||
175 GTTC 172

```

```

seq_name: gb_est2:BG760879

```

```

seq_documentation_block:

```

```

LOCUS      BG760879      598 bp      mRNA      linear      EST 15-MAY-2001
DEFINITION 602717063F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4840886 5',
            mRNA sequence.
ACCESSION  BG760879
VERSION    BG760879.1  GI:14071532
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 598)
            NIH-MGC http://mhc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: ATCC/DCPD/DRP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LHCMI672 row: m column: 15
            High quality sequence stop: 596.
            Location/Qualifiers
              1. 598
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4840886"
                /clone_lib="NIH_MGC_49"
                /tissue_type="melanotic melanoma, high MDR (cell line)"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: skin; Vector: pORF7; Site: 1: XhoI; Site: 2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GCCACAG(G). Size-selected >500bp for average insert size
                1.8kb. Library constructed by Ling Hong in the laboratory
                of Gerald M. Rubin (University of California, Berkeley)
                using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                II RT (Life Technologies). Note: this is a NIH_MGC
                library."

```

```

BASE COUNT      153 a      160 c      166 g      119 t
ORIGIN

```

```

alignment_scores:
  Quality: 130.50      Length: 168
  Ratio: 1.466        Gaps: 6
  Percent Similarity: 52.976      Percent Identity: 26.190

```

```

alignment_block:
US-09-701-711-2 x BG760879 ..

```

```

Align seg 1/1 to: BG760879 from: 1 to: 598

```

```

651 GUAasnPhetYrAlaGlyTYrGlySerValArgGlyTYrAspGlnse 667
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
29 GATAGGTTTAACTTGGGGACCACCAAGCATCCGCGATTACAGATGCA 78
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGln 684
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
79 CAGCATGGGGCCACAGACCA..... 100
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
684 InThrThrLeuGlyGlnValAlaGlyAlaLeuAlaThrPheGly 700
      |||||  |||||  |||||  |||||  |||||  |||||  |||||
101 .....GGAGACTACCTAGTGGAGAACGCTAGTGGCGCGCGC 139
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
701 SerGlnLeuLeuProLeuProPheLys.....GlyAspTrrpII 714
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
140 CTGACCTCTACACCCCATTCCTTCGGCAGGCCAGCGGTGGCTTGG 189
      :::::|||||  |||||  |||||  |||||  |||||  |||||  |||||
714 eaSpGlnValArgProValIlePheIleGlyGlyGlnValPheAsp 731

```


438 CAGGGGGGCGACAGATTGTGATGTGTCCAGTTTGAGCTGGGATTTCG 487
 812 lPhe 813
 ||||
 488 GTC 491

seq_name: gb_est1:BB471006

seq_documentation_block:

LOCUS BB471006 661 bp mRNA linear EST 25-OCT-2001
 DEFINITION BB471006 RIKEN full-length enriched, 12 days embryo eyeball Mus
 musculus cDNA clone D23003624 3', mRNA sequence.

ACCESSION BB471006

VERSION BB471006.2 GI:16427864

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 661)

AUTHORS

Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
 Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,
 M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
 Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,
 D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
 Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
 Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
 Unpublished (2001)
 On Jul 22, 2000 this sequence version replaced gi:3388195.

TITLE

Journal Contact: Yoshihide Hayashizaki

COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216

Email: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
 M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)

Wagl,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
 Watanabe,M., Itoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsunura
 S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and
 Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.
 10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
 Y. and Hayashizaki,Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
 K., Fukuda,S., Hara,A., Itoh,M., Kawai,T., Shibata,K. and
 Hayashizaki,Y.
 Computational Analysis of Full-length Mouse cDNAs Compared with
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp/) for
 further details.

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

FEATURES

source

1. 661
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="D23003624"

/clone.lib="RIKEN full-length enriched, 12 days embryo
 eyeball"
 /tissue_type="eyeball"
 /dev_stage="12 days embryo"
 /lab_host="DH10B"
 /note="Site_1: Sali; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5']
 GAGGAGAGAGCGCGCCGCAACCTGAGTTTCTTTTCTTNN 3']. cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. Second strand cDNA was prepared with the
 primer adapter of sequence [5']
 GAGGAGAGATTCTCGAGTGAATTAATTAATTAATTCCTCCCTCC 3']. cDNA
 was cleaved with BamHI and XhoI. Vector: a modified
 pBluescript KS(+) after bulk excision from Lambda FLX I."

BASE COUNT 135 a 179 c 194 g 153 t

ORIGIN

alignment_scores:
 Quality: 130 50 Length: 168
 Ratio: 1.450 Gaps: 6
 Percent Similarity: 53.571 Percent Identity: 26.786

alignment_block:

US-09-701-711-2 x BB471006

Align seg 1/1 to: BB471006 from: 1 to: 661

651 glnasnphetryalaglytyrelservalarglytyrpsglnse 667
 :::::::::::::::::::: |:::::::::::::::::::
 122 GACAGGTTTACCTGCGAGGCCGCCAGAGTGCAGAGATTATGATGCA 171
 :::::::::::::::::::: |:::::::::::::::::::
 667 rserleuglyproargserglnalatyrlleuthralargarglygln 684
 :::::::::::::::::::: |:::::::::::::::::::
 172 CAGCATTTGACCCGACAGAGTGA..... 193
 684 lnrtrthleuglyglvalvalaglyglasnalaleualatrphlegly 700
 :::::::::::::::::::: |:::::::::::::::::::
 194GGAGATTACTGCGCGCGGCGGAGCCTACTGGCTGGGGC 232
 701 sergluleuileuoproleuprophelys.....glyasptrip1 714
 :::::::::::::::::::: |:::::::::::::::::::
 233 CTGCACCTCTACACCCCTGCTCCGCGCAGCGGCGGTGCTCGG 282
 714 easpglnvalargprovaliilepheiiegluglyglvalpheaapt 731
 :::::::::::::::::::: |:::::::::::::::::::
 283 AGAGCTTTTCAGAACTCCTTTTCCTCAATGCGGGCAACCTG..... 325
 731 nrtrnglymetasplysglntrhrlleaspleuthrghlphelyaspro 747
 :::::::::::::::::::: |:::::::::::::::::::
 326TGCAACCTCACTATGATGAGAGGCGCC 352
 748 glnalatrhalagluglnasnalalysalalaasnarargproleueth 764
 :::::::::::::::::::: |:::::::::::::::::::
 353 AAAGCCCATATCCGGAACCTGATGAGTGC..... 382
 764 rglaspslysglnleuargtyrseralaglvalglvalatrtrp...T 780
 :::::::::::::::::::: |:::::::::::::::::::
 383ATCCGCTGCTCTATGAGAGCGGCTGCTCCGAC 419
 780 yrtrhrproleuglyproleuserllesertryalalysproleuasnlys 796
 :::::::::::::::::::: |:::::::::::::::::::
 420 TTGGCAACATGCTCGCTGCGTGCAGTGCATGCTATGAGGCGCTG 469
 797 lysglinsnaspsgln...Thraspthrvalglnpheglniieglyserva 812
 :::::::::::::::::::: |:::::::::::::::::::
 470 CAGGGGGGCGACAGATTGTGATGTGTCCAGTTTGAGCTGGGATTTCG 519

```

812 1phe 813
520 GTTC 523

seq_name: gb_est2:BE612912

seq_documentation_block:
LOCUS BE612912 759 bp mRNA linear EST 20-OCT-2000
DEFINITION 60145184T1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3855778 3',
mRNA sequence.
ACCESSION BE612912
VERSION BE612912.1 GI:9894590
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE NIH-MGC http://mgi.mcg.mcg.gov/
AUTHORS 1 (bases 1 to 759)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs@femail.nih.gov
Tissue Procurement: DCM/DMP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM9583 row: 9 column: 11
High quality sequence start: 38
High quality sequence stop: 751.
Location/Qualifiers
1. 759
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="3855778"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-Sport6; Site: 1: Not;
Site: 2: Salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 174 a 210 c 199 g 176 t
ORIGIN
alignment_scores:
Quality: 130.50 Length: 168
Ratio: 1.466 Gaps: 6
Percent Similarity: 52.976 Percent Identity: 26.190
alignment_block:
US-09-701-711-2 x BE612912/rev
Align seg 1/1 to reverse of: BE612912 from: 1 to: 759
651 GUAaNPheTYrAlaGlyTYrGlySerValaTYrAspGlnse 667
:::||||| ||||| ||||| ||||| |||||
540 GATAGGTTTACCTTGGGGGACCAACACATCCGGGATTCAGCATGCA 491
667 rSerleuGlyProaGrgSerGlnAlaTYrleuThralaArGaTgGlyGlnG 684
||||| ||||| ||||| ||||| |||||
490 CAGCAATCGGGCCACAGAGCAA..... 469
684 InThrThrlcuGlyGlnValaGlyGlyAsnaAlaLeuAlaThrpheGly 700
||||| ||||| ||||| ||||| |||||
468 .....GAGACTACCTCAGGTGGAGAGAGCTACTGGGCGGCGGC 430
701 SerGlnleuLeuLeuProleuProPhelys.....GlyAspTrpI 714

```

```

429 CTGACACCTCTACACCCCTTACCTTCCGCCAGGCCAGGGGTGCTTGG 380
714 eaSpGlnValaArgProValaIlePheIleGlyGlyGlnValaPheAsp 731
:::||||| ||||| ||||| ||||| |||||
379 AGAAGCTTTCCGACACACTCTCTTCCACAGCAGGAACCTC..... 337
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheIleAspPro 747
336 .....TGCAACCTCAACTATGGGAGGCGCC 310
748 GlnAlaThrAlaGlnAlaGlnAlaValaAlaAlaAsnaArgProleuLeuTh 764
309 AAAGCTCATATTCGTAGCTGCTGAGTGC..... 280
764 rGlnAspLysGlnLeuArgTYrSerAlaGlyValaGlyAlaThrTrp...T 780
279 .....ATCCGCTGTCGTACGGGGCGCGGATGTCCTCAGGC 243
780 YrThrProIleGlyProleuSerIleSerTYrAlaLysProleuAsnLys 796
242 TTGGCAACATCGCTCGGTGGAACTTATTAATGCGTCCCATGGAGTA 193
192 CAGACAGCGCACAGATATGATGCGCTCCAGTTGGAGCTGGATAG 143
812 1phe 813
142 GTTC 139

seq_name: gb_gss:BH391826
seq_documentation_block:
LOCUS BH391826 788 bp DNA linear GSS 11-DEC-2001
DEFINITION AG-ND-17815.TF ND-TAM Anopheles gambiae genomic clone AG-ND-17815,
DNA sequence.
ACCESSION BH391826
VERSION BH391826.1 GI:17337967
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anophelinae.
1 (bases 1 to 788)
REFERENCE Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
AUTHORS Direct Submission of BAC-end sequences from Anopheles gambiae
TITLE Unpublished (2001)
JOURNAL Other_GSSs: AG-ND-17815.TF
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TrGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae
to minimize the inclusion of DNA from microorganisms that inhabit
the gut. The DNA is derived from mixed sexes of larvae. The BAC
library was constructed at Texas A&M University BAC Center
University, College Station, Texas 77843-2123, USA using a HindIII
partial digest.
Seq primer: M13 For
Class: BAC ends.
Location/Qualifiers
1. 788
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"

```

FEATURES

source

```

/clone="AG-ND-17815"
/clone_lib="ND-TAM"
/notes="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      214 a      150 c      119 g      305 t
ORIGIN

alignment_scores:
  Quality: 130.50      Length: 267
  Ratio: 0.826      Gaps: 11
  Percent Similarity: 59.176      Percent Identity: 21.723

alignment_block:
US-09-701-711-2 x BH391826/rev ..

Align seg 1/1 to reverse of: BH391826 from: 1 to: 788

38  ILeThrGlyLeuGlnArgValThrIleGlySerLeuGlnSerValLeuPr 54
    :::::::::::::::::::: ||| |||:::
785 GTGATGCTTAAAAAATACACCTGACCAATTTAGCTTACAGCG 736
    :::::::::::::::::::: ||| |||:::
54  oPheArgLeuGlyGlnValSer.....GluAsnGlnLeuAlaAspG 69
    :::::::::::::::::::: ||| |||:::
735 TCTGTGAAGAAGTGAATCTGTGAATCCCGGCAAAAGACTGAGTACAG 686
    :::::::::::::::::::: ||| |||:::
69  lyVallyAlaLeuTyraIleThrGlyAsnPheserAspValGlnValTy 85
    :::::::::::::::::::: ||| |||:::
685 CTATCAAAAAGCTTTGGGATAGCCAAATATTTCTGAGTTAAAGTATAT 636
    :::::::::::::::::::: ||| |||:::
86  HisGln.....GluGlyArgIleIle.....TyrGlnValThrG 97
    :::::::::::::::::::: ||| |||:::
635 GTGACAGCATTTGAGACAAATATTCCTTAAATTCCTCTTACAGA 586
    :::::::::::::::::::: ||| |||:::
97  uArgProLeuIleAlaGluIleAsnPhesGlyGlyAsnArgLeuIlePro 114
    :::::::::::::::::::: ||| |||:::
585 TTAAAGAAAGCTGTGTAAGTAAATTTACCGGA...AAGGAATCAAAA 539
    :::::::::::::::::::: ||| |||:::
114 yAsGlyLeuGlnGlyLeuLysAsnLagIleuAlaVal.GlyG 130
    :::::::::::::::::::: ||| |||:::
538 AATCTAAAAAGAAAATTTATTAAGATTAATTAAGCTGAAGCTGGTAT 489
    :::::::::::::::::::: ||| |||:::
130 nProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsn 146
    :::::::::::::::::::: ||| |||:::
488 GAATATTACCGACCGAGAACTGTTACTAATCACTAATGATGACCTCAG 439
    :::::::::::::::::::: ||| |||:::
147 GlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrVal..... 161
    :::::::::::::::::::: ||| |||:::
438 CAATACATTAACAAGAGATCCCGATGCTAAGATCACCATTTGAAGATA 389
    :::::::::::::::::::: ||| |||:::
161 sGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPheAlaG 178
    :::::::::::::::::::: ||| |||:::
388 AATCAATGCAAGAGATCTAGCTTATGACTGACCATTAATTAAGTACATA 339
    :::::::::::::::::::: ||| |||:::
178 LngIlyAspProAlaArgValAlaAspIleAsnIleIleGlyAsnGlnHis 194
    :::::::::::::::::::: ||| |||:::
338 AAGCAAGACCTGTTAAGATGACCGTATGATTTTGAAGAAACAACAGT 289
    :::::::::::::::::::: ||| |||:::
195 PheserAspAlaAspLeuIleAspValLeuAlaIleLysAspAsnLys 210
    :::::::::::::::::::: ||| |||:::
288 GTTTCCTCACTAAACT...CGTAAAGCGTTTAAATAATGCAAAACA 242
    :::::::::::::::::::: ||| |||:::
211 .....IleAsnProLeuSerLysAlaAspArgTyrThrGlnGlu 224
    :::::::::::::::::::: ||| |||:::
241 AAGAGATTCCTCTGTTGCTTAAAGCTTGCAAAATTTATCAACAGATA 192
    :::::::::::::::::::: ||| |||:::
191 AATATGAGAAGATTAAGACATCTTGTAGACTACTAACAATCTTTAGGA 142
    :::::::::::::::::::: ||| |||:::
241 PheValArgPheGluIleLysAspAlaLysLeu.....AsnIleAs 254
    :::::::::::::::::::: ||| |||:::
141 TTT.....AGAGATATGAGAGTAGTGTCAAGACTCTGTTTC 107
    :::::::::::::::::::: ||| |||:::

```

```

254  nGluAspLysAsnArgIlePheValGlnIleSerLeuHisGluGlyGluG 271
    :::::::::::::::::::: ||| |||:::
106 GCGTAATATCTAAGCATCAACATCAAAAGTAAATAGATGAGGTAATAA 57
    :::::::::::::::::::: ||| |||:::
271  InTyArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThr 286
    :::::::::::::::::::: ||| |||:::
56  AATACTATATCGGAGATATTTACATTCTGAGCAAAATACAGTATTTTCT 10
    :::::::::::::::::::: ||| |||:::

seq_name: gb_est1:AM107426

seq_documentation_block:
LOCUS      AM107426                871 bp      mRNA      linear      EST 20-Oct-1999
DEFINITION un16c03.x1 Sugano mouse kidney mKia Mus musculus cDNA clone
IMAGE:2192452.3' similar to WP:034E10.1 CE01181 ;, mRNA sequence.
ACCESSION  AM107426
VERSION    AM107426.1  GI:6078226
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 871)
AUTHORS   Maira,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
            The WashU-NCI Mouse EST Project 1999
            Unpublished (1999)
            Other_ESTs: un16c03.y1
            Contact: Maira M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@wustl.edu
            This clone is available royalty-free through LBNL; contact the
            IMAG Consortium (info@image.llnl.gov) for further information.
            MGI:1004904
            Seq primer: custom primer used
            High quality sequence stop: 461.
            Location/Qualifiers
                1..871
                /organism="Mus musculus"
                /strain="C57BL"
                /db_xref="taxon:10090"
                /clone="IMAGE:2192452"
                /clone_lib="Sugano mouse kidney mKia"
                /sex="female"
                /dev_stage="adult"
                /lab_host="DH10B"
                /note="Organ: Kidney; Vector: pME18S-FL3; Site_1: DraIII
                (CACTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                was primed with an oligo(dT) primer
                (ATGTCGCGCTTTTCTTTTCTTTT); double-stranded cDNA was
                ligated to a DraIII adaptor [TGTGGCCCTAGCG], digested
                and cloned into distinct DraIII sites of the pME18S-FL3
                vector (3' site CACTGTG, 3' site CACCATGTG). XhoI should
                be used to isolate the cDNA insert. Size selection was
                performed to exclude fragments <1.5kb. Library
                constructed by Dr. Sumio Sugano (University of Tokyo
                Institute of Medical Science). Custom primers for
                sequencing: 5' end primer CTCTGCTCTTAAAGCTGCG and 3' end
                primer CGACTGCGCTGAGCAACA."

BASE COUNT      201 a      254 c      227 g      181 t      8 others
ORIGIN

alignment_scores:
  Quality: 130.50      Length: 168
  Ratio: 1.450      Gaps: 6
  Percent Similarity: 53.571      Percent Identity: 26.786

```

alignment block:

US-09-701-711-2 x AM107426/rev ..

Align seg 1/1 to reverse of: AM107426 from: 1 to: 871

```

651  GLUAsnpheTYrAlaGlyTYrGlySerValArGgLYTYrAspGlnse 667
      ::::::::::::::::::::|::::::::::::::::::
538  GACAGGTTTACCTGGAGGCCACAGAGTGTCCGAGGATTACCATCA 489
      ::::::::::::::::::::|::::::::::::::::::
667  rSerLeuGlyProArGSerGlnAlaTYrLeuThrAlaArGArgLYGlnG 684
      ::::::::::::::::::::|::::::::::::::::::
488  CAGATTGGACCCACAGAGTAA..... 467
      ::::::::::::::::::::|::::::::::::::::::
684  InThrThrLeuGlyGluValAlaGlyGlnAlaLeuAlaThrPheGly 700
      ::::::::::::::::::::|::::::::::::::::::
466  .....GGAGATTACCTGGCGCGCGGAGGCTACTGGCGTGGGGGC 428
      ::::::::::::::::::::|::::::::::::::::::
701  SerGluLeuLeuProLeuProPheLYs.....GlyAspTPil 714
      ::::::::::::::::::::|::::::::::::::::::
427  CTGCACCTCTACACCCACCTGCCCTTCGGCCAGCCAGGCGGCTTCGG 378
      ::::::::::::::::::::|::::::::::::::::::
714  eAspGluValArProValIlePheIleGluGlyGlyGlnValPheAsp 731
      ::::::::::::::::::::|::::::::::::::::::
377  AGACCTTTTCAGAACTCATTTCCTCAATGCGGCAACCTG..... 335
      ::::::::::::::::::::|::::::::::::::::::
731  hrThrGlyMetAspLYsGlnThrIleAspLeuThrGlnPheLYsAspPro 747
      ::::::::::::::::::::|::::::::::::::::::
334  .....TGCNACTCAACTATGTTGGTGGGGGCC 308
      ::::::::::::::::::::|::::::::::::::::::
748  GlnAlaThrAlaGlnAlaGlnAlaLYsAlaAlaAsnArProLeuLeuTh 764
      ::::::::::::::::::::|::::::::::::::::::
307  AAGCCCATATCCGAGAGCTAGCTAGTGC..... 278
      ::::::::::::::::::::|::::::::::::::::::
764  rGlnAspLYsGlnLeuArGTYrSerAlaGlyValGlyAlaThrTP...T 780
      ::::::::::::::::::::|::::::::::::::::::
277  .....ATCCGCTGTGCTATGAGCAGCGCTGCTCCGAC 241
      ::::::::::::::::::::|::::::::::::::::::
780  YrThrProIleGlyProLeuSerIleSerTYrAlaLYsProLeuAsnLYs 796
      ::::::::::::::::::::|::::::::::::::::::
240  TTGGCAACATCCGCTGCGTGCAGCTGACACTGCACTTCCTATGGCGGTG 191
      ::::::::::::::::::::|::::::::::::::::::
797  LYsGlnAsnAspGln...ThrasPthrValGlnPheGlnIleGlySerVa 812
      ::::::::::::::::::::|::::::::::::::::::
190  CAGGGGGGGCAGACGATTGTGTGATGCTGCAGTTGGAGCTGGAGTTCG 141
      ::::::::::::::::::::|::::::::::::::::::
812  lPhe 813
      ::::::::::::::::::::|::::::::::::::::::
140  GTTC 137

```

seq_name: gb_est2:BM233588

seq_documentation_block:

```

LOCUS      BM233588          580 bp      mRNA      linear      EST 31-JAN-2002
DEFINITION K0338H12-3 NIA Mouse Osteoblast cDNA Library (Long) Mus musculus
ACCESSION  BM233588
VERSION    BM233588.1 GI:17868858
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 580)
AUTHORS   Plao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G.,
Umezawa, A. and Ko, M.S.H.
TITLE     Systematic Analyses of NIA Mouse Osteoblast cDNA Library (Long)
JOURNAL   Unpublished (2001)
COMMENT   Contact: Dawood B. Dudekula
          Laboratory of Genetics
          National Institute on Aging/National Institutes of Health
          333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
          Email: cdnaelg@nigms.nih.gov
          Plate: K0338 row: H column: 12

```

Seq primer: -21M13 Forward
High quality sequence stop: 580
POLY-A-Tes.

FEATURES

source

Location/Qualifiers

1..580
/organism="Mus musculus"
/strain="C3H/He mice"
/db_xref="nla:EST:K0338H12-3"
/db_xref="taxon:10090"
/clone="K0338H12"
/clone_lib="NIA Mouse Osteoblast cDNA Library (Long)"
/tissue_type="Osteoblast"
/cell_line="KUSA-A1 cells"
/lab_host="DH10B"
/note="vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were obtained from Dr. Akihito Umezawa (Keio University School of Medicine, Japan). Double-stranded cDNAs were synthesized with an Oligo(dt) primer [Invitrogen: 5'-DCACTAGTTCTAGATCGGAGCGCGCCCTTTT-3'] from 2.1 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lone-linker L-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Tag polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 3.0 kb. The library was constructed by Yulan Plao (NIA)."

BASE COUNT 131 a 174 c 157 g 118 t
ORIGIN

alignment_scores:

Quality: 129.50 Length: 168
Ratio: 1.439 Gaps: 6
Percent Similarity: 53.571 Percent Identity: 26.786

alignment block:

US-09-701-711-2 x BM233588/rev ..

Align seg 1/1 to reverse of: BM233588 from: 1 to: 580

```

651  GLUAsnpheTYrAlaGlyTYrGlySerValArGgLYTYrAspGlnse 667
      ::::::::::::::::::::|::::::::::::::::::
540  GACAGGTTTACCTGGAGGCCACAGAGTGTCCGAGGATTACCATCA 491
      ::::::::::::::::::::|::::::::::::::::::
667  rSerLeuGlyProArGSerGlnAlaTYrLeuThrAlaArGArgLYGlnG 684
      ::::::::::::::::::::|::::::::::::::::::
490  CAGATTGGACCCACAGAGTAA..... 469
      ::::::::::::::::::::|::::::::::::::::::
684  InThrThrLeuGlyGluValAlaGlyGlnAlaLeuAlaThrPheGly 700
      ::::::::::::::::::::|::::::::::::::::::
468  .....GGAGATTACCTGGCGCGCGGAGGCTACTGGCGTGGGGGC 430
      ::::::::::::::::::::|::::::::::::::::::
701  SerGluLeuLeuProLeuProPheLYs.....GlyAspTPil 714
      ::::::::::::::::::::|::::::::::::::::::
429  CTGCACCTCTACACCCACCTGCCCTTCGGCCAGCCAGGCGTGCCTTCG 380
      ::::::::::::::::::::|::::::::::::::::::
714  eAspGluValArProValIlePheIleGluGlyGlyGlnValPheAsp 731
      ::::::::::::::::::::|::::::::::::::::::
379  AGACCTTTTCAGAACTCATTTCCTCAATGCGGCAACCTG..... 337
      ::::::::::::::::::::|::::::::::::::::::
731  hrThrGlyMetAspLYsGlnThrIleAspLeuThrGlnPheLYsAspPro 747
      ::::::::::::::::::::|::::::::::::::::::

```

```

336 ..... TGCACCTCAACTATGGTGGAGGGCCC 310
748 GlnAla1ThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
309 AAGGCCATATCCGAGACGACTGAGTGC..... 280
764 rglnAspLysGlnLeuArgTyrSerIleArgLysAlaGlyAlaThrTrp...T 780
279 .....ATCCGCTGGTCTATGAGAGAGGCGGTGTCCTCCGAC 243
780 yThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
242 TTGGCAGCATTCGCTGGCTGGACCTGAACACACGATTCCTATGGGCGTG 193
797 LysGlnAsnAspGln...ThraspThrValGlnPheGlnIleGlySerVa 812
192 CAGGGGGGGACAGCATTTGTGTGATGTGTCCACTTTGAGACTGGGATTCG 143
812 lphe 813
142 GTTC 139
seq_name: gb_est2:BI277013
seq_documentation_block:
LOCUS      BI277013              534 bp      mRNA      linear      EST 19-JUL-2001
DEFINITION  UI-R-CY0-bxp-b-10-0-UI-s1 UI-R-CY0 Rattus norvegicus cDNA clone
ACCESSION  UI-R-CY0-bxp-b-10-0-UI 3', mRNA sequence.
VERSION    BI277013
KEYWORDS   BI277013.1 GI:14922480
SOURCE     EST.
ORGANISM   Norway rat.
            Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 534)
AUTHORS    Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
COMMENT    Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            non-normalized brown adipose library cDNA library Preparation: M.B
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=yes.
FEATURES             location/Qualifiers
     source           1..534
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CY0-bxp-b-10-0-UI"
                     /clone_11b="UI-R-CY0"
                     /dev_stage="ABULR"
                     /lab_host="DHI0B (Life Technologies)"
                     /note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CY0
library is a non-normalized library constructed from rat
brown adipose tissue. For a detailed description of the
library from which this clone was derived, please visit

```

our web site at ratstat.eng.uiowa.edu. The subtraction has been previously described in (Donald, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_L1B-UI-R-CY0
TAG_TISSUE-brown adipose
TAG_SEQ-TTCGC*

BASE COUNT 118 a 165 c 149 g 101 t 1 others

ORIGIN

alignment_scores:
Quality: 125.50 Length: 168
Ratio: 1.410 Gaps: 6
Percent Similarity: 52.976 Percent Identity: 26.190

alignment_block:
US-09-701-711-2 x B1277013/rev ..

Align seg 1/1 to reverse of: B1277013 from: 1 to: 534

```
651 GUAAsnPhetYrAlaGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      :::::|||||  |||  |||||||||||:::
509 GACAGGTTTACTCTGAGAGGCCCCACGACGAGCGTCGAGATTAGCATGCA 460
      :::::|||||:::
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 664
      :::::|||||:::
459 CACACATCGGACCCAGAGCGAA..... 438
684 InThrLeuGlyGlyAlaValAlaGlyAsnAlaLeuAlaThrPheGly 700
      ||:::  :::::|||||:::  |||
437 .....GGAGATTACTGTGGCGGGGAGGACATACGTGGCGGGGCGC 399
      :::::|||||  |||||||||||:::
701 SerGlnLeuLeuLeuProLeuProPheLys.....GlyAspTrrP11 714
      :::::|||||  |||||||||||:::  |||  :::
398 CTCACCTCTATACCCACCTGCTTCCGCGCAGCCAGCGAGGTGGCTTCG 349
714 eAspGlnValArgProValIlePheIleGlyGlyGlnValPheAsp 731
      :::::  |||  :::::|||||:::
348 AGAGCTTTTCAGAACTCAGCTTTTCTCATAGCTGGCAACCTG..... 306
731 hrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspPro 747
305 .....TGCACCTCACTACGAGGTGAGGCCCC 279
748 GlnAlaThrAlaGlnGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      :::::  :::::|||||:::
278 AGAGCCACATCCGGAAGCTGGCTGATGC..... 249
764 rGlnAspLysGlnLeuArgTyrSerAlaGlyAlaGlyAlaThrTrp...T 780
248 .....ATCCGCTGTGCTATGCTGAGGACATCGCTCCGCGC 212
780 yThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLys 796
211 TTGGCAACAATGCTCGCTGAGCTAACAACACTGCATTCTATAGGGGTG 162
797 LysGlnAsnAspGln...ThrAspThrValGlnPheGlnIleLysSerVa 812
      :::::|||||  |||  |||||||
161 CAGCGAGGCGACAGATTTGTGTATGCGCTCCACATTGGAGCTGGATTGCG 112
812 lphe 813
      |||
111 GTTC 108
```

seq_name: gb_est2:BE647176

seq_documentation_block:
LOCUS BE647176 541 bp mRNA linear EST 06-SEP-2000
DEFINITION UI-M-BH1-ame-c-09-0-UI.r1 NIH.BMP.M.S2 Mus musculus cDNA clone
ACCESSION BE647176
VERSION BE647176.1 GI:9972958
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 541)
TITLE Bonaldo,M.F., Lennon,G. and Soares,M.B.
JOURNAL Normalization and subtraction: two approaches to facilitate gene discovery
MEDLINE Genome Res. 6 (9), 791-806 (1996)
COMMENT 97044477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: MEST@mail.nih.gov
CDNA Library Preparation: M.B. Soares Lab Clone distribution:
Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M3 Reverse

FEATURES
Source
Location/Qualifiers
1..541
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH1-ame-c-09-0-UT"
/clone_1lb="NIH-BMAP_M_S2"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="vector: pUT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH-BMAP_M_S2 library is a subtracted library derived from NIH-BMAP_M_S1, which in turn is a subtracted library derived from a mixture of normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus). The driver used for subtraction consisted of a pool of 5,000 clones from the NIH-BMAP_M_S1 library and a pool of 2,000 clones obtained from non-normalized and normalized mouse brain spinal cord libraries."
BASE COUNT 106 a 148 c 162 g 125 t
ORIGIN

alignment_scores:
Quality: 124.50 Length: 168
Ratio: 1.383 Caps: 6
Percent Similarity: 53.571 Percent Identity: 26.190

alignment_block:
US-09-701-711-2 x BE647176 ..

Align seg 1/1 to: BE647176 from: 1 to: 541

```

651 GUAAnPhETyAlAGlyTyrglySerValArGlyTyTAspGlnSe 667
      ::::::::::::::::::::
132 GACAGGTTTACCTGGAGGCCCCACGAGGTGTCGAGGATTTCATGCA 181
      ::::::::::::::::::::
667 rSerLeuGlyProArGSerGlnAlaTyTLeuThrAlaArGArGlyGlnG 684
      ::::::::::::::::::::
182 CAGCATTTGACCCCGAGAGTGA..... 203
      ::::::::::::::::::::
684 InThrThLeuGlyGluValAlGlyGlyAsnAlaLeuAlaThrPhGly 700
      ::::::::::::::::::::
204 .....GGAATTAACCTGGCGCGGAGGCTTACTGGGCTGGGCGC 242
      ::::::::::::::::::::
701 SerGluLeuIleLeuProLeuProPhelys.....GlyAspTrpI 714
      ::::::::::::::::::::

```

```

243 CTCACCTCTACACCCACCTCCCGGCCAGGCCAGGGTGCTCGG 292
714 eAspGlnValArProValIlePheIleGluGlyGlnValPheAspT 731
      ::::::::::::::::::::
293 AGACCTTTTCAGACCTCATTTTCCATTCGCGGCAACCTG..... 335
731 hrThrGlyMetAspGlyGlnThrIleAspLeuThrGlnPheAspPro 747
336 .....TCGACCTCAACTGATGGTGAGGGCCCGC 362
748 GlnAlaThrAlaGluGlnAsnAlaValAlaAsnArProLeuLeuTh 764
363 AAGCCCATTCGCGAGACCTGAGACTGC..... 392
764 rGlnAspGlyGlnLeuArGTrSerAlaGlyValAlaThrTrp...T 780
393 .....ATCGGTGTCCTATGAGAGAGGCTGCTCCGAC 429
780 YrThrProIleGlyProLeuSerIleSerThrAlaLysProLeuAsnLys 796
430 TTGGCAACATCGCTCGGCTGAGCTGAACTGCACTTCCTATGCGCTG 479
797 LysGlnAsnAspGln...ThAspThrValGlnPheGlnIleGlySerVa 812
480 CAGGGGGGCGACAGATTGTGTATGGTGTCCACTTTGGAGCTGGATTGC 529
812 lPhe 813
530 GTTC 533
seq_name: gb_est1:AW027570

```

seg documentation_block:
LOCUS AW027570 596 bp mRNA linear EST 09-MAR-2000
DEFINITION w74h01.x1 Soares_thymus.NHFrth Homo sapiens cDNA clone
IMAGE:2535313 3' similar to SW:YDF6.SCH0 Q10478 HYPOTHETICAL 51.8
KD PROTEIN C173.06 IN CHROMOSOME 1.; mRNA sequence.
ACCESSION AW027570 GI:5886326
VERSION
KEYWORDS
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 596)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps@email.nih.gov
NOTE: This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert length: 1344 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 457.
FEATURES
source
Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2535313"
/clone_1lb="Soares_thymus.NHFrth"
/dev_stage="fetal"
/lab_host="DH10B (Phage-resistant)"
/note="Organ: thymus, pooled; Vector: pUT3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
TGTTACCAATCTGAGTGGAGCGCGCCGACGATTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pUT3 vector. Library
went through one round of normalization. Library

constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 124 a 171 c 165 g 136 t

ORIGIN

alignment_scores: Quality: 124.50 Length: 168
Ratio: 1.415 Gaps: 6
Percent Similarity: 52.381 Percent Identity: 26.190

alignment_block:
US-09-701-711-2 x AM027570/rev ..

Align seg 1/1 to reverse of: AM027570 from: 1 to: 596

```

651 GUASNPHEYYTALAGLYTYRGLYSERVALARGLYTYRASPGLNSE 667
      ::::::::::::::::::::|
554 GATAGGTTTACCTCGGGGACCCACAGCGTCCGGATTCCAGCATGCA 505
      |||:|||||:|||||:
667 rSerLeuGLYProArGserGlnAlaTYrLeuThraAlaArgAGLYGNG 684
      |||:|||||:|||||:
504 CAGCATCGGGCCACAGCGGAA..... 483
      |||:|||||:|||||:
684 InThrThrLeuGLYGLYValAGLYGlnAlaLeuAlaThrpheGLY 700
      |||:|||||:|||||:
482 .....GGAGACTACCTAGGTGAGAGAGCTACTGGGCGCGGCGC 444
      ::::::::::::::::::::|
701 SerGluLeuIleuProLeuProPhelys.....GLYAspTrpIi 714
      ::::::::::::::::::::|
443 CTGCACCCCTTACACCCCACTTACCTTCCGGCCAGCGCCAGGCGCTTGC 394
      |||:|||||:|||||:
714 eAspGlnValArgProValIlePheIleGLYGLYValAGLYValaThrp 731
      ::::::::::::::::::::|
393 AGAAGCTTTCCGACACACACTTCTTCTCAGCGAGAAACCTC..... 351
      |||:|||||:|||||:
731 hrThrGLYMetAspLysGLInThrIleAspLeuThrGlnPhelysAspPro 747
      ::::::::::::::::::::|
350 .....TGCAACCTCACTATGCGGAGGAGGCGCC 324
      |||:|||||:|||||:
748 GlnAlaThraIAGlnGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      ::::::::::::::::::::|
323 AAAGCTCATATTCTGAAGCTGCTGAGTGC..... 294
      |||:|||||:|||||:
764 rGlnAspLysGlnLeuArgTYrSerAlaGLYValAGLYValaThrp...T 780
      ::::::::::::::::::::|
293 .....ATCCGCTGCTGACGGGCGCGGATGTCCTCAGGC 257
      |||:|||||:|||||:
780 YrThrProIleGLYProLeuSerIleSerTYrAlaLysProLeuAsnLys 796
      |||:|||||:|||||:
256 TTGGCAACATCGCTCGTTGGACTTATCTGCGTCCCATGGAGTA 207
      ::::::::::::::::::::|
797 LysGlnAsnAspGln...ThraSPThrValAGlnPheGlnIleGLYSerVa 812
      ::::::::::::::::::::|
206 CAGCAGAGTGCAGGATATGTATGTGCGCTCCAGTTGGAGCTGGGATAG 157
      |||:|||||:|||||:
812 Lphe 813
      |||:|||||:|||||:
156 GTTC 153

```

seq_name: gb_est1:AM65916

seq_documentation_block:

LOCUS AM65916 598 bp mRNA linear EST 06-APR-2000
DEFINITION h195f07.x1 Soares_NFL_T_GBC_SI Homo sapiens cDNA clone
IMAGE:2980069.3' similar to SW:YDF8_SCHPO Q10478 HYPOTHETICAL 51.8
KD PROTEIN C17C9.06 in CHROMOSOME 1. ; mRNA sequence.
ACCESSION AM65916
VERSION AM65916.1 GI:7458465
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 598)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
COMMENT Email: cgaps-remail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: -400P from Gluco
High quality sequence stop: 429.
Location/Qualifiers

FEATURES
source

```

1..598
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2980069"
/clone_lib="Soares_NFL_T_GBC_SI"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7/3D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NDHL19W, testis NHT, and B-cell
NCI-CGAP_GCB1) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 125 a 170 c 163 g 140 t
ORIGIN

```

alignment_scores: Quality: 123.50 Length: 168
Ratio: 1.403 Gaps: 6
Percent Similarity: 52.381 Percent Identity: 26.190

alignment_block:

US-09-701-711-2 x AM65916/rev ..

Align seg 1/1 to reverse of: AM65916 from: 1 to: 598

```

651 GUASNPHEYYTALAGLYTYRGLYSERVALARGLYTYRASPGLNSE 667
      ::::::::::::::::::::|
556 GATAGGTTTACCTCGGGTACCCACAGCGTCCGGGATTCAGCATGCA 507
      |||:|||||:|||||:
667 rSerLeuGLYProArGserGlnAlaTYrLeuThraAlaArgAGLYGNG 684
      |||:|||||:|||||:
506 CAGCATCGGGCCACAGCGGAA..... 485
      |||:|||||:|||||:
684 InThrThrLeuGLYGLYValAGLYGlnAlaLeuAlaThrpheGLY 700
      |||:|||||:|||||:
484 .....GGAGACTACCTAGGTGAGAGAGCTACTGGGCGCGCGC 446
      ::::::::::::::::::::|
701 SerGluLeuIleuProLeuProPhelys.....GLYAspTrpIi 714
      ::::::::::::::::::::|
445 CTGCACCTTACACCCCACTTACCTTCCGGCCAGCGCCAGGCGCTTGG 396
      |||:|||||:|||||:
714 eAspGlnValArgProValIlePheIleGLYGLYValAGLYValaThrp 731
      ::::::::::::::::::::|
395 AGAAGCTTTCCGACACACACTTCTTCTCAGCGAGAAACCTC..... 353
      |||:|||||:|||||:
731 hrThrGLYMetAspLysGLInThrIleAspLeuThrGlnPhelysAspPro 747
      ::::::::::::::::::::|
352 .....TGCAACCTCACTATGCGGAGGAGGAGGCGCC 326
      |||:|||||:|||||:
748 GlnAlaThraIAGlnGlnAsnAlaLysAlaAlaAsnArgProLeuLeuTh 764
      ::::::::::::::::::::|
325 AAAGCTCATATTCTGAAGCTGCTGAGTGC..... 296
      |||:|||||:|||||:
764 rGlnAspLysGlnLeuArgTYrSerAlaGLYValAGLYValaThrp...T 780

```


JOURNAL
COMMENT

Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 414.
 Location/Qualifiers

FEATURES
source

1..556
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2629806"
 /clone_lib="NCI-CCAP_Ut1"
 /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 111 a 159 c 155 g 130 t 1 others
 ORIGIN

alignment_scores:
 Quality: 120.50 Length: 168
 Ratio: 1.369 Gaps: 6
 Percent Similarity: 52.381 Percent Identity: 26.190

alignment_block:
 US-09-701-711-2 x AM150700/rev ..

Align seg 1/1 to reverse of: AM150700 from: 1 to: 556

```

651 GluAsnPhetYralaglyTYrglySerValarglyTYraspglnse 667
      :::::|||||
556 GATAGGTTTAACCTCGGGGAGCCCAAGCGTCCGCGATTGACATGCA 507
      :::::|||||
667 rSerleuglyProargSerGlnalatyrlleuthAlaargglygling 684
      :::::|||||
506 CAGCATCGGGCCACAGACGAA..... 485
684 lnThrThrleuglyGluValaIglylnaAlaAlaThrPhagly 700
      :::::|||||
484 .....GGAGACTACTAGTGGAGAGCGTANTGGCGCGCGC 446
701 SerGluLeulleuProleuProPhelys.....GlyAspTrpIl 714
      :::::|||||
445 CTGCACCTTAACACCCCATTCCTCCGGCCAGCCAGGTGGCTTGG 396
714 eaapGlnValargProvalIlephelleglyglyGlnValpheaspr 731
      :::::|||||
395 AGAAGCTTTCCGAACACACTTCTTCAACGCGGAACCTC..... 353
731 hrThrGlymetAsplysGlnThrIleaspleuthrGlnPhelyaspPro 747
      :::::|||||
352 .....TGCAACTCAACTATGGGAGAGGCCCC 326
748 GlAlaThrAlaGluGlnaAlaLysAlaAlaAsnArgProleuLeuth 764
      :::::|||||
325 AAAGCTCATTTGCTAGTGGCTGACTGC..... 296
764 rGlnAsplysGlnleuArgTyrSerAlaGlyValaIAlaThrTrp...T 780
      :::::|||||
295 .....ATCCGCTGTGCTAGCGGGCGCGGATGTCTCTCAAGC 259

```

```

780 yrThrProIleglyProleuserIleSerTYraIaLysProleuAsnLys 796
      :::::|||||
258 TTGGCAACATCGCTCGGTTGGACTTAATTACTGCTCCCAATGGAGTA 209
797 LysGlnaAspGln...ThrAspThrValGlnPheginIleglySerVa 812
      :::::|||||
208 CAGACAGGTGACAGATATGTGATGCGCTCCAGTTGGAGCTGGGATAG 159
812 lPhe 813
      :::::|||||
158 GTTC 155

```


Align seg 1/1 to: AAZ29550 from: 1 to: 2442

```
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17
|||||
1 ATCGGTAATTCATATTTAAAGGTTTCAGGTCAGTCAATGGC 50
|||||
17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAla 34
|||||
51 TGTCAATGATGTAATGCAACTCATGCACAAAGCGGGGATTTTATGGCAA 100
|||||
34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50
|||||
101 ATCACATTACCATCAGAGGACTACAGGAGTGACCATTTGAAGCTTACAA 150
|||||
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67
|||||
151 AGCGTGCTGCCGTTTCGCTGGGTCAAGTGGTGAGCGAAGAACAGTTGGC 200
|||||
67 aspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84
|||||
201 TGTATGTTGTCAAAGCACTTTATGCAACAGGCAATTTTTCAGATGTGCAAG 250
|||||
84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100
|||||
251 TCTATCATCAAGAAGGCGGTATCATCTATCAGGTAACCGAAGGCCGTTA 300
|||||
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117
|||||
301 ATCGCTGAGATTAATTTTGAGGCGCAATCGCTTAATTCCAAAGAAAGTCT 350
|||||
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134
|||||
351 ACAAGAAGGCGTAAAAAATGCTGGCTTAGCTGGGTCAACCACTAAAC 400
|||||
134 lnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150
|||||
401 AAGCCACAGTACAGATGATCGAAGCCGAGCTTACCAATCAATATATATCA 450
|||||
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGl 167
|||||
451 CAAGGCTATTATTAATACCGAAATTAAGTCAAAAGGTAACCTGCTTGGT 500
|||||
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184
|||||
501 TAATCGTTGTAAGCTTGATGATGACCTTTGCTGAAGGTAAACCTGCAAGG 550
|||||
184 aValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
|||||
551 TGGTTGATTAATATATCATTTGGCAATCAGCATTTTAGCGATGCAATTTG 600
|||||
201 IleAspValLeuAlaIleLysaspAsnLysIleAsnProLeuSerLysAl 217
|||||
601 ATTGATGCTTGGCGATTAAAGGATATAAAATCAATCCACTGCTCTAAAGC 650
|||||
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234
|||||
651 TGACCGTTTATCTCAGAAAGCTGGTGACCATTTAGAGAAATTTGCGGTG 700
|||||
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysaspAlaLys 250
|||||
701 CTAATATATCTCAATGCAGGTTTGTGCGTTTGGAGATTAAGATGCTAAG 750
|||||
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267
|||||
751 CTTAATATTAAATGAAGATAAAACCGTATCTTTTGTGTGAGATTTCAATTGCA 800
|||||
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
|||||
801 TGAAGGTGACCAATATCCGTTTGGACAGACACAGTTTTCGGTAAATTTAA 850
|||||
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
|||||
851 CTTATACTCAAGCAGAACTTGAGGCACCTGCTTAAATTCAAAGCAGAAGAA 900
|||||
```

```
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
|||||
901 GGGTTTTTCACAAGCCATGCTTGAGCAACAACAACAATATCATGACCAA 950
|||||
317 sPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334
|||||
951 ATTGTGTGACGATGGCTATTATTATGCTCAATCCGTCCTGTAAACAGCA 1000
|||||
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
|||||
1001 TTAATGATGAAAGTCGTACGGTTGATGTGGAATATTATATGACCCGTGA 1050
|||||
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGl 367
|||||
1051 CACCCGTCTATGTAGTACGCCGTATTATTTTACAGGTAACCTTAAGACCCA 1100
|||||
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384
|||||
1101 AGATGAAGTACTCCGTCGTGAGATGCGAACCTTGAAGGTGCGTGGCAT 1150
|||||
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
|||||
1151 CTAATCAAAAATCCAGCTGCTCGTGACGCTTGATGCGGACTGGGTTT 1200
|||||
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspGl 417
|||||
1201 TTATAACATGTTACCGTTGATCTCGTCAGTACCACCACTCACCTCATCA 1250
|||||
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434
|||||
1251 GGTGATGTAAATTTTGTGTTGAGAAACAACCTTCAGGATCATCAACCA 1300
|||||
434 leAlaAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspVal 450
|||||
1301 TCGCAGCAGGCTACTCTCAAAGTGGTGGTAACTTTTCAATTTGATGTT 1350
|||||
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
|||||
1351 TCTCAAAATAACTTTATGGGTACAGTAAGCAGTCAATGCTTCGTTTTC 1400
|||||
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
|||||
1401 TCGCTCTGAGACCCGCTGAGGTGTATAGTTTGGGTATGACCAACCCATACT 1450
|||||
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
|||||
1451 TTACCGTAATGGCGTCTCGCAAAGCTTGAGTGGCTACTATCGCAAAACC 1500
|||||
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGl 517
|||||
1501 AAGTATGATAACAACAAGAACATTAGTAATTTATGTTACTTGTATCTTATGGTGG 1550
|||||
517 ySerLeuSerTyrGlyTyrProIleaspGluAsnGlnArgIleSerPheG 534
|||||
1551 CTCATTAGCTATGGATATCCAAATTCGAATCAACGCAATGAAGCTTTG 1600
|||||
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
|||||
1601 GTCTGAATGCTCACAAATACCAAGCTTCATGGCGGTCGTTTATGGGCATT 1650
|||||
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
|||||
1651 AGTAATGTCAAGCAGCTGATGCGAGATGGTGCAAAATTCAAAGTGGATAA 1700
|||||
567 nAsnGlyIleProAspPheLysHisAspTyrThrTyrAsnAlaIleL 584
|||||
1701 TAATGGCATTCCTGATTTTAAGCATGATTACACACCTTCAATGCCATTT 1750
|||||
584 euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
|||||
1751 TGGGGTGGAAATTTATTCAAGTCTAGATCGCCCTGTATTTCCAACCCCAAGGC 1800
|||||
```

601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisG1 617
|||||
1801 ATGAGTCATCTGTAGATTGACGGTTGGTTTGGTGATAAACTCATCA 1850

617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysServ 634
|||||
1851 AAAAGTGGTTTATCAAGGCAATATCTATCGCCCATTTATCAAAAAATCAG 1900

634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnLeuProPheTyr 650
|||||
1901 TCTTCGGTGGATACCCCAAGTAGCTATGGCAATAATTACCATTTTAT 1950

651 GluAsnPheTyrAlaGlycylTyrGlySerValArgGlyTyrAspGlnse 667
|||||
1951 GAAAAATTTCTATGACAGCGCGGTATGTTGCGGTTCGTGGCTATGATCAATC 2000

667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||
2001 CTCCTTGGGTCACGCTCACAGCCATATTGACAGCTCGTCGTGTCAC 2050

684 InThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
|||||
2051 AAACACACTAGGACAGGTGTTGGTGGTATGCTTTGGCAACTTTCGCG 2100

701 SerGluLeuIleLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717
|||||
2101 ACTGAGCTGATTTTACCTTTGCGCATTTAAAGGTGATTGGATAGATCAGGT 2150

717 IargProValIlePheIleGluGlycylGlnValPheAspThrThrGlyM 734
|||||
2151 CGGTCCAGTGATATTCAATTGAGGGCGGTGAGGTTTGTATACACAGGTA 2200

734 eAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
|||||
2201 TGGATAACAACCACTGATTGTTAAACCAATTAAAGACCCACCAACA 2250

751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLy 767
|||||
2251 GCTGAACAAAATGCAAAAGCAGCCCAATCGCCGCTACTAACCAAGATA 2300

767 scGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784
|||||
2301 ACAGTTGCGGTATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2350

784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
|||||
2351 GTCCCTTTATCTATTAGCTATGCCAGCCATTGAATAAAAAACAATAATGAT 2400

801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
2401 CAGACCGATACGGTACAGTTCAGATTGCGTAGTGTCTTT 2439

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAZ29551

seq_documentation_block:

ID AAZ29551 standard; DNA; 2442 BP.

XX AC AAZ29551;

XX AC AAZ29551;

XX DT 14-MAR-2000 (first entry)

XX DE M. catarrhalis (ATCC 43617) BASB027 polypeptide encoding DNA #2.

XX KW BASB027; OMP85; outer membrane protein; otitis media; diagnosis;
XX KW treatment; screening; bacterial infection; ds.

XX OS Moraxella catarrhalis.

XX PH Key Location/Qualifiers

XX FT CDS 1..2442

XX FT /*tag= a
XX FT /product= "BASB027 polypeptide" -

PN WO9963093-A2.
XX
PD 09-DEC-1999.
XX
PF 31-MAY-1999; 99WO-EP03822.
XX
PR 03-JUN-1998; 98GB-0011945.
XX
PR 08-MAR-1999; 99GB-0005304.
XX
PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Vinals-Bassols C;
XX
DR WPI; 2000-105700/09.
XX
DR P-PSDB; AAY44391.
XX
PT Novel BASB027 polynucleotide and polypeptides from Moraxella
XX catarrhalis useful for treating M. catarrhalis infection such as otitis
XX media
XX
PS Claim 8; Page 102; 109pp; English.
XX
CC The present sequence is a DNA obtained from chromosomal DNA library of
XX Moraxella catarrhalis strain Mc2931 (ATCC 43617). It encodes BASB027
XX polypeptide, which shows significant homology to Neisseria meningitidis
XX OMP85 outer membrane protein. BASB027 polynucleotide and polypeptide can
XX be used for diagnosis and staging of disease, determining susceptibility
XX to a disease and to prepare medicaments for treating M. catarrhalis
XX infections, especially otitis media. The BASB027 DNA can be used as
XX probe for screening of genetic mutations, serotype, taxonomic
XX classification or identification. BASB027 agonists, antagonists and
XX antibodies may be used to prevent and/or treat bacterial infections.
XX
SQ Sequence 2442 BP; 735 A; 461 C; 535 G; 711 T; 0 other;

alignment_scores:

Quality: 749.00 Length: 749

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAZ29551

Align seg 1/1 to: AAZ29551 from: 1 to: 2442

65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAs 81

|||||
193 CAGTTGGCTGATGGTCAAGCACTTTATGCAACAGGCAATTTTTCAGA 242

81 pValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGlu 98

|||||
243 TGTGCAAGTCTATCATCAAGAAGGGCGGTATCATCTATCAGGTAACCGAAA 292

98 rgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLys 114

|||||
293 GSCCGCTTAATCGCTGAGATTAATTTTGGGGCAATCGCTTAATTCACAAA 342

115 GluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131

|||||
343 GAAGGCTACAAAGAGGGCTAAAAAATGCTGGCTTAGCTGTGGGTCAACC 392

131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148

|||||
393 ACTAAAAAGCCACAGACATGATCGAAACCGAGCTTACCAATCAATCAAT 442

148 yTleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMet 164

|||||
443 ATATATCACAGGCTATTATATACCGAAATTACTGTCAACAGAGGATG 492

165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181

|||||
493 CTTGATGGTAATCGTGTAAAGCTTGATATGACCTTTGCTGGAAGGTAACC 542

|||||
2343 GCCCATTGGCTCTTTATCTATTAGCTATGCCAAGCCATTGAATAAAAAAC 2392

798 InAsnAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||

2393 AAATGATGACAGCGATACGGTACAGTTCCAGATGGTGTCTTT 2439
|||||

seq_name: /SDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF28550

seq_documentation_block:

ID AAF28550 standard; DNA; 99629 BP.

XX AAF28550;

XX 04-APR-2001 (first entry)

XX Genomic fragment #37.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.

XX Moraxella catarrhalis.

XX WO200078968-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.

XX (INCY-) INCYTE GENOMICS INC.

XX Lagace RE, Patterson C, Berg KL;

XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -

XX Claim 1; Page 391-415; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.

XX SQ Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other;

alignment_scores:

Quality:	749.00	Length:	749
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-701-711-2 x AAF28550 ..

Align seg 1/1 to: AAF28550 from: 1 to: 99629

65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAs 81
|||||

66325 CAGTTGGCTGATGGTCAAGCACTTATGCAACAGGCAATTTTCAGA 66374
|||||

81 pValcInValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluA 98
|||||

|||||
66375 TGTCCAAGTCTCATCAAGAAGGGCGGTATCATCTATCAGTAACCGAAA 66424
|||||

98 rgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLys 114
|||||

66425 GGCCGTTAATCCCTGAGATTAAATTTTGGGGCAATCGCTTAATTCACAAA 66474
|||||

115 GluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131
|||||

66475 GAAGTCTACAAAGAGGGCTAAAAAATGCTGGCTTAGCTGGGTCAACC 66524
|||||

131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148
|||||

66525 ACTAAACCAAGCCACAGTACAGATCGAAACCCAGCTTACCAATCAAT 66574
|||||

148 yrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMet 164
|||||

66575 ATATATCACAGGCTTATTAATACCGAAATTAAGTCAACACAGCATG 66624
|||||

165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181
|||||

66625 CTTGATGGTAATCGTTAAGCTTGATATGACCTTTGCTGAAGGTAAACC 66674
|||||

181 oAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspA 198
|||||

66675 TGCACGGGTGGTTGATATTAATATCATTTGGCAATCAGCATTTTAGCGATG 66724
|||||

198 laAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsnProLeu 214
|||||

66725 CAGATTTGATTGATGCTTCGGATTAAAGGATATAAAATCAATCCACTG 66774
|||||

215 SerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAs 231
|||||

66775 TCTAAAGCTGACCGTTATACTCAAGAAAAGCTGGTGACCACTTTAGAGAA 66824
|||||

231 nLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysA 248
|||||

66825 TTTGCGTGCTAAATATCTCAATGCAAGGTTTGTGCGTTTTCAGATTAAAG 66874
|||||

248 spAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIle 264
|||||

66875 ATGCTAAGCTTAATTAATAAGATAAAACCGTATCTTTGTGTGAGATT 66924
|||||

265 SerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGl 281
|||||

66925 TCATTGTCATGAAGTGAGCAATATCGCTTTGGCAGACACACAGTTTTCGG 66974
|||||

281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLysPheLysA 298
|||||

66975 TAATTTAACTTATACTCAAGCAGAACTTGAGGCACCTGCTTAAATTCAAA 67024
|||||

298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314
|||||

67025 CAGAAGAAGGGTTTTCACAGCCATGCTTGAGCAAAACAAACAATAATC 67074
|||||

315 SerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProVa 331
|||||

67075 AGTACCAAAATTTGGTGAGCATGGCTATTATTATGCTCAAAATCCGTCCTGT 67124
|||||

331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348
|||||

67125 AACACGCATTAATGATGAAGTCGTACGGTTGATCTGGAATATTATATTG 67174
|||||

348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364
|||||

67175 ACCCTGTACACCCCTGTATGTACCGCGTATTAATTTTACAGGTAACCTT 67224
|||||

365 LysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381
|||||

67225 AAGACCAAGATGAAGTACTCCGTGCGTACGATCGCACAACTTTGAAGGTGC 67274
|||||

381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398
|||||

67275 GTTGGCATCTAATCAAAAAATCAGCTGCTGTCACGCTTGATGCGGA 67324
 398 hrGlyPhePheLysHisValThrValAspThrArgProValProAsnSer 414
 67325 CTGGGTTTTTAAACAATGTTACCGTGTGATACGCTGATACCAACTCA 67374
 415 ProAspGlnValAspValAsnPheValValGluGlnProSerGlySe 431
 67375 CCTGATCAGTTGATGATAAATTTGTGGTTGAAGAACACCTTCAGGATC 67424
 431 rSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnP 448
 67425 ATCAACCATCGCAGCAGGCTACTCTCAAAAGTGGTGGTGAACCTTTTCAAT 67474
 448 heAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAla 464
 67475 TTGATGTTTTCTCAAAATAACTTTATGGGTACAGGTAAAGCAGCTCAATGCT 67524
 465 SerPheSerArgSerGlnThrArgGluValTyrSerLeuGlyMetThrAs 481
 67525 TCGTTTTCTCGCTCTGAGACCCGCTGAGGTGTATAGTTTGGGTATGACCAA 67574
 481 nProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrA 498
 67575 CCCATACCTTTACCGTAAATGGCGCTCTCGCAAGCTTGAGTGGCTACTATC 67624
 498 rGlySerThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSer 514
 67625 GTAAACCAAGATATGATAACAAGACATTTAGTAATATGCTACTTGATTCT 67674
 515 TyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIle 531
 67675 TATGTTGGCTCATTTAGCTATGATATCAATTCATGATAAATCAACGCAT 67724
 531 eSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheM 548
 67725 AGCTTTGCTCTGAATGCTGCAATACCAAGCTTCAATGGCGGTGCTTTTA 67774
 548 etGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGln 564
 67775 TGGGCATTAGTAATGTCAAGCAGCTGATGCCAGATGGTGGCAAAATTCAA 67824
 565 ValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrTyrAs 581
 67825 GTGGATAATAATGGCATTCCTGATTTAAGCATGATTACACAACCTTACAA 67874
 581 nAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgProValPheProT 598
 67875 TGCCATTTTGGGGTGGGAATATTCAAGTCTAGATCGCCCTGTATTCCAA 67924
 598 hrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAspLys 614
 67925 CCCAAGCATGATCATCTCTAGATTTTACGCGTTGGTTTGGTGATAAA 67974
 615 ThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleIly 631
 67975 ACTCATCAAAAAGTGGTTTATCAAGCAATATCTATCGCCCAATTTATCAA 68024
 631 sLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuP 648
 68025 AAAATCAGTCTTGGGTGGATAGCCCAAGTTAGGCTATGGCAATAATTTAC 68074
 648 roPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyr 664
 68075 CATTTTATCAAAATTTCTATGACGGCGGTATGGTTTCGGTTCGTGCTAT 68124
 665 AspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgAr 681
 68125 GATCAATCTCTTTGGTCCAGCCTCACAAAGCCTATTTCACAGCTCGTGC 68174
 681 gGlyGlnGlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaT 698
 68175 TGGTCAACAAACACACTAGAGAGAGTGTGGTGGTAATGCTTTGGCAA 68224

698 hrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTrpIle 714
 68225 CTTTCGGCAGTGAAGCTGATTTTACCTTTGCCATTTAAAGTGATGGATA 68274
 715 AspGlnValArgProValIlePheIleGluGlyGlyGlnValPheAspTh 731
 68275 GATCAGTCCGTCAGTGATATTCAATGAGGCGGTGAGGTTTTCATAC 68324
 731 rThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProG 748
 68325 AACAGGTATGGATAACAACCACTGATTAAACCAATTTAAAGACCCAC 68374
 748 InAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThr 764
 68375 AAGCAACAGCTGAACAAATGCAAAAGCAGCAATGCCCGCTACTAAC 68424
 765 GlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrTh 781
 68425 CAAGATAAACAGTTGCGTTATAGTGTGTTGGTGCAACTGGTATAC 68474
 781 rProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysG 798
 68475 GCCCATGTGCTCTTATCTATTAGCTATGCCAAGCCRTTGAATATAAAAC 68524
 798 InAsnAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
 68525 AAAATGATCAGACCGATACGGTACAGTTCAGATTGGTAGTGCTTT 68571
 seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF91415
 seq_documentation_block:
 ID AAF91415 standard; DNA; 1000 BP.
 XX
 AC AAF91415;
 XX
 DT 04-MAY-2001 (first entry)
 XX
 DE Moraxella catarrhalis D15 gene upstream sequence, SEQ ID:41.
 XX
 KW Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
 KW genetically modified; protective antigen expression; LPS detoxification;
 KW LPS; lipid A; homologous recombination vector; immunisation;
 KW immunoprotective; non-toxic; paediatric; ds.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200109350-A2.
 XX
 PD 08-FEB-2001.
 XX
 PF 31-JUL-2000; 2000WO-EP07424.
 XX
 PR 03-AUG-1999; 99GB-0018319.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Berthet FJ, Dalemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;
 PI Poolman J, Thily G, Thonnard J, Voet P;
 XX
 XX WPI; 2001-138654/14.
 XX
 PT New isolated polynucleotide useful for outer membrane vesicle
 PT preparation from Gram-negative bacterial strain for vaccination of
 PT microbial infections -
 XX
 PS Claim 46; Page 87; 128pp; English.
 XX
 CC The invention relates to a genetically-engineered outer membrane vesicle
 CC (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
 CC The blebs of the invention are improved with respect to their
 CC immunogenicity and toxicity by the introduction of one or more genetic
 CC changes to the chromosome of the bacterium from which the blebs are

CC derived. The changes made include the upregulation of protective antigen
CC expression, the downregulation of immunodominant non-protective antigen
CC expression, and genetic changes which result in detoxification of the
CC Lipid A moiety of lipopolysaccharide (LPS). The invention also
CC encompasses modified gram-negative bacterial strains from which the bleb
CC preparations are made, a vector suitable for performing recombination
CC events (for the generation of the modified bacterial strains),
CC bacterially-derived nucleic acid sequences used in such a vector, and an
CC immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
CC cell vaccine suitable for paediatric use. The bleb preparation is useful
CC in the manufacture of a medicament for immunising a human host against a
CC disease caused by infection of one or more of the following: *Neisseria*
CC meningitidis, *Neisseria gonorrhoeae*, *Haemophilus influenza*, *Moraxella*
CC catarrhalis, *Pseudomonas aeruginosa*, *Chlamydia trachomatis*, and *Chlamydia*
CC pneumonia. The invention may also be used to provide immunisation against
CC the influenza virus. Bacterially derived nucleotide sequences of the
CC invention are used in the performance of homologous recombination events
CC up to 1000 bp upstream of a bacterial chromosomal gene in order to either
CC increase or decrease expression of that gene. Immunoprotective and
CC non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines
CC are more immunogenic, less toxic and safer, and are particularly useful
CC for paediatric use. The present sequence represents a specifically
CC claimed *Moraxella catarrhalis* nucleic acid sequence.
XX
SQ Sequence 1000 BP; 287 A; 172 C; 215 G; 326 T; 0 other;

alignment_scores:
Quality: 36.00 Length: 36
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAF91415 ..

Align seg 1/1 to: AAF91415 from: 1 to: 1000

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAla 17
|||||
821 ATGGCTAATTCATATTTTAAAGGTTTTTCAGGTCAGTGCATGACAATGGC 870

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAla 34
|||||
871 TGTGATGATGGTAATGTCAACTCATGTCACAAAGCGCGGATTATGGCAA 920

34 snAspIle 36
|||||

921 ATGACATT 928

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ29554

seq_documentation_block:

ID AAZ29554 standard; DNA; 61 BP.

AC AAZ29554;

DT 14-MAR-2000 (first entry)

DE M. catarrhalis BASB027 amplifying forward PCR primer MC-D15-BamF.

XX PCR primer; MC-D15-BamF; BASB027; otitis media; OMP85;

KW outer membrane protein; treatment; diagnosis; bacterial infection; ss.

XX Synthetic.

OS *Moraxella catarrhalis*.

XX WO963093-A2.

PN 09-DEC-1999.

PD 31-MAY-1999; 99WO-EP03822.

XX 03-JUN-1998; 98GB-0011945.

PR 08-MAR-1999; 99GB-0005304.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Vinals-Bassols C;

XX WPI; 2000-105700/09.

XX Novel BASB027 polynucleotide and polypeptides from *Moraxella*
PT catarrhalis useful for treating *M. catarrhalis* infection such as otitis
PT media -

XX Example 2; Page 51; 109pp; English.

XX The present sequence is a forward PCR primer MC-D15-BamF used for
CC amplifying chromosomal DNA extracted from 16 *Moraxella catarrhalis*
CC strains for variably analysis of the BASB027 gene.

XX Sequence 61 BP; 20 A; 15 C; 17 G; 9 T; 0 other;

alignment_scores:

Quality: 11.00 Length: 11
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAZ29554 ..

Align seg 1/1 to: AAZ29554 from: 1 to: 61

39 ThrGlyLeuGlnArgValThrIleGluSerLeu 49
|||||
28 ACAGGACTACAGGAGTGACCATTCGAAGCTTA 60

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:AAZ29555

seq_documentation_block:

ID AAZ29555 standard; DNA; 67 BP.

AC AAZ29555;

DT 14-MAR-2000 (first entry)

DE M. catarrhalis BASB027 amplifying reverse PCR primer MC-D15-SalRC.

XX PCR primer; MC-D15-SalRC; BASB027; otitis media; OMP85;

KW outer membrane protein; detection; treatment; bacterial infection; ss.

XX Synthetic.

OS *Moraxella catarrhalis*.

XX WO963093-A2.

XX 09-DEC-1999.

XX 31-MAY-1999; 99WO-EP03822.

XX 03-JUN-1998; 98GB-0011945.

XX 08-MAR-1999; 99GB-0005304.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Vinals-Bassols C;

XX WPI; 2000-105700/09.

XX Novel BASB027 polynucleotide and polypeptides from *Moraxella*
PT catarrhalis useful for treating *M. catarrhalis* infection such as otitis
PT media -

XX Example 2; Page 51; 109pp; English.

XX

CC The present sequence is a reverse complementary PCR primer MC-D15-SalRC
 CC used for amplifying chromosomal DNA extracted from 16 Moraxella
 CC catarrhalis strains for variability analysis of the BASB027 gene.
 XX
 SQ Sequence 67 BP; 22 A; 16 C; 15 G; 14 T; 0 other;

alignment_scores:
 Quality: 11.00 Length: 11
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-701-711-2 x AAZ29555/rev ..

Align seg 1/1 to reverse of: AAZ29555 from: 1 to: 67

803 AspThrValGlnPheGlnIleGlySerValpHe 813
 |||||
 66 GATACGTCAGTCCAGATTGATGTCCTTT 34

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAF60966

seq_documentation_block:
 ID AAF60966 standard; DNA; 1830 BP.

XX AAF60966;

XX 16-MAY-2001 (first entry)

DE P. putida KT2440-associated DNA ORF00612.

XX Transgenic plant; detection; probe; amplification; vaccine carrier;
 KW microbial production strain; biological remediation; ds.

OS Pseudomonas putida.

XX DE19935088-A1.

PD 01-FEB-2001.

PF 27-JUL-1999; 99DE-1035088.

PR 27-JUL-1999; 99DE-1035088.

XX (TIGR-) TIGR INST GENOMIC RES.

PA (QUIA-) QUIAGEN GMBH.

PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.

PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.

XX WPI; 2001-192469/20.

XX New DNA sequences specific for Pseudomonas putida KT2440, useful as
 PT safe genetic engineering host, allow detection in presence of other
 PT related bacteria -

XX Claim 1a; Page 13-14; 158pp; German.

XX This invention describes novel DNA sequences (I) for specific detection
 CC of Pseudomonas putida KT2440. The invention also describes (1)
 CC recombinant expression vector containing (1); (2) prokaryotic or
 CC eukaryotic cells transformed or transfected with (1) or the vector of
 CC (1); (3) production of expression products by culturing cells of (2);
 CC (4) expression products, or their fragments, of (1) and synthetic
 CC proteins or peptides with the same sequences (A); (5) poly- or
 CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
 CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
 CC plants that contain transformed or transfected cells of (2); (8)
 CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
 CC carrying one or more (1). (I), and their fragments, are used as probes
 CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
 CC polymerase chain reaction, and for production of transgenic plants. (I),

CC or antibodies that recognize their expression products, are used for
 CC detecting the presence of KT2440, particularly in presence of other,
 CC even closely related, bacteria. KT2440 is one of the bacteria classified
 CC as safe, by the National Institutes of Health, for genetic engineering
 CC work, e.g. as microbial production strains, for biological remediation
 CC and as vaccine carriers. (I) are exclusive to KT2440 with no significant
 CC homology with sequences in other bacteria (specifically the closely
 CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
 CC has greater catabolic activity and better survival in, and adaptation to,
 CC the rhizosphere and soil.
 XX
 SQ Sequence 1830 BP; 401 A; 555 C; 529 G; 344 T; 1 other;

alignment_scores:
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAF60966 ..

Align seg 1/1 to: AAF60966 from: 1 to: 1830

368 AspGluValLeuArgArgGluMetArgGln 377

|||||

1081 GACGAAGTGTCTGCGTCCGAAATCGCCAG 1110

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.AAS85853

seq_documentation_block:

ID AAS85853 standard; cDNA; 3750 BP.

XX AAS85853;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #21657.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG21666.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

PS Claim 1; SEQ ID No 21657; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS84197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3750 BP; 937 A; 928 C; 965 G; 920 T; 0 other;

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAS85853 ..

Align seg 1/1 to: AAS85853 from: 1 to: 3750

648 ProPheTyrGluAsnPhetYrAlaGlyGly 657
|||||
1435 CGGTTCTACGAGACTTCTATGCCGTGGT 1464

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:AAT84180

seq_documentation_block:

ID AAT84180 standard; DNA; 1000 BP.

XX AAT84180;

DT 14-SEP-1998 (first entry)

DE DNA encoding a Staphylococcus aureus protein of unknown function.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW Staphylococcal gene; regulatory element; bacterial gene expression;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW toxic shock syndrome; ss.

XX Staphylococcus aureus.

OS Key Location/Qualifiers
FH CDS complement (357..647)
FT /*tag= a

XX WO9730070-A1.

XX 21-AUG-1997.

XX 19-FEB-1997; 97WO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI; 1997-424969/39.

DR P-PSDB; AAW28281.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT to isolate antimicrobial compounds, and in vaccines against S.

PT aureus infection

XX Claim 9; Page 940-941; 989pp; English.

XX The present sequence encodes a Staphylococcus aureus protein of unknown
CC function. The present sequence was isolated from a library of clones of
CC S. aureus WCUH 29 in Escherichia coli. The DNA sequence can be used in
CC the construction of ribozymes and antisense sequences to control the
CC expression of Staphylococcal genes. The DNA sequence is also useful
CC as a source of regulatory elements for the control of bacterial gene
CC expression. The encoded protein may be used to produce vaccines to
CC enable a host to produce specific antibodies with antibacterial action.
CC These vaccines and antibodies would protect a host against invasion by
CC S. aureus, and conditions relating to Staphylococcal infection, e.g.
CC Staphylococcal food poisoning, scaled skin syndrome, and toxic shock
CC syndrome.

XX SQ Sequence 1000 BP; 289 A; 235 C; 151 G; 313 T; 12 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAT84180/rev ..

Align seg 1/1 to reverse of: AAT84180 from: 1 to: 1000

717 ValArgProValIlePheIleGly 725
|||||
311 GTGCGACCCGTCATTTTATTGAAGGA 285

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:AAV53520

seq_documentation_block:

ID AAV53520 standard; DNA; 1000 BP.

XX AAV53520;

XX 30-OCT-1998 (first entry)

DE DNA encoding a Exou protein.

XX Staphylococcus aureus protein; immune response induction; eye infection;
KW antibody production; i-cell immune response; gastrointestinal infection;
KW respiratory infection; inhibitor; bacterial infection; cardiac infection;
KW central nervous system; kidney infection; urinary tract infection;
KW antimicrobial compound identification; broad spectrum antibiotic;
KW therapy; ss.

XX Staphylococcus aureus.

XX EP841394-A2.

XX 13-MAY-1998.

XX 24-SEP-1997; 97EP-0307485.

XX 24-SEP-1996; 96US-0027032.

XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Black MT, Burnham MKR, Hodgson JE, Knowles DJC;
PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M;
PI Ward JM;

XX WPI; 1998-252940/23.

DR P-PSDB; AAW77731.

XX New nucleic acid sequences from Staphylococcus aureus WCUH29 -

PT useful in vaccines and for treatment of bacterial infections of e.g.
PT respiratory tract and central nervous system
XX
PS Claim 1; Page 185; 390pp; English.
XX
CC This sequence encodes a *Staphylococcus aureus* protein that (based on
CC homology with a *Rhizobium meliloti* protein) is a
CC Exon protein, and represents a DNA sequence of the invention.
CC The DNA sequences were isolated from *Staphylococcus aureus* WCHU29
CC (NCIMB 40771). Host cells containing the DNA sequences are used to
CC produce polypeptides or fragments. The proteins are used in the treatment
CC of disease, for inducing an immune response by administering them, to
CC produce antibody and/or T-cell immune response. Antagonists of the
CC proteins are used for the inhibition of bacterial polypeptides.
CC Conditions which may be treated include bacterial infections, especially
CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,
CC urinary tract, skin, bones and joints. The proteins can also be used to
CC identify antimicrobial compounds which are broad spectrum antibiotics,
CC especially useful in the treatment of *H. pylori* infection.
XX
SQ Sequence 1000 BP; 289 A; 235 C; 151 G; 313 T; 12 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAV53520/rev ..

Align seg 1/1 to reverse of: AAV53520 from: 1 to: 1000

717 ValArgProValIlePheIleGluGly 725
|||||
311 GTGCGACCCGTCATTTTATTGAAGGA 285

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAF61065

seq_documentation_block:

ID AAF61065 standard; DNA; 2019 BP.
XX
AC AAF61065;
XX
DT 16-MAY-2001 (first entry)
XX
DE *P. putida* KT2440-associated DNA ORF08500.
XX
KW Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
OS *Pseudomonas putida*.
XX
PN DE19935088-A1.
XX
PD 01-FEB-2001.
XX
PF 27-JUL-1999; 99DE-1035088.
XX
PR 27-JUL-1999; 99DE-1035088.
XX
PA (TIGR-) TIGR INST GENOMIC RES.
PA (QUIA-) QUIAGEN GMBH.
PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
DR WPI; 2001-192469/20.
XX
PT New DNA sequences specific for *Pseudomonas putida* KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT related bacteria -
XX

PS Claim 1a; Page 117-118; 158pp; German.

XX
CC This invention describes novel DNA sequences (I) for specific detection
CC of *Pseudomonas putida* KT2440. The invention also describes (1)
CC recombinant expression vector containing (I); (2) prokaryotic or
CC eukaryotic cells transformed or transfected with (I) or the vector of
CC (1); (3) production of expression products by culturing cells of (2);
CC (4) expression products, or their fragments, of (I) and synthetic
CC proteins or peptides with the same sequences (A); (5) poly- or
CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
CC plants that contain transformed or transfected cells of (2); (8)
CC detecting KT2440 using a labeled (I) or Ab as probe; and (9) DNA chips
CC carrying one or more (I). (I), and their fragments, are used as probes
CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
CC polymerase chain reaction, and for production of transgenic plants. (I),
CC or antibodies that recognize their expression products, are used for
CC detecting the presence of KT2440, particularly in presence of other,
CC even closely related, bacteria. KT2440 is one of the bacteria classified
CC as safe, by the National Institutes of Health, for genetic engineering
CC work, e.g. as microbial production strains, for biological remediation
CC and as vaccine carriers. (I) are exclusive to KT2440 with no significant
CC homology with sequences in other bacteria (specifically the closely
CC related pathogen *P. aeruginosa*). Compared with other 'safe' bacteria, it
CC has greater catabolic activity and better survival in, and adaptation to,
CC the rhizosphere and soil.
XX
SQ Sequence 2019 BP; 404 A; 671 C; 603 G; 338 T; 3 other;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAF61065 ..

Align seg 1/1 to: AAF61065 from: 1 to: 2019

368 AspGluValLeuArgArgGluMetArg 376
|||||
250 GACGAAGTGTCTGCGCGAAATGCGC 276

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAL29004

seq_documentation_block:

ID AAL29004 standard; DNA; 50 BP.
XX
AC AAL29004;
XX
DT 24-JAN-2002 (first entry)
XX
DE Human SNP oligonucleotide #2212.
XX
KW Immunosuppressive; immunostimulatory; antiinflammatory; cytostatic;
KW neuroprotective; antimicrobial; gene therapy; vaccine; amylase; cancer;
KW amyloid protein; angiotensin; apoptosis related protein; cadherin;
KW cyclin; polymerase; oncogene; histone; kinase; colony stimulating factor;
KW complement related protein; cytochrome; kinesin; cytokine; interferon;
KW interleukin; G-protein coupled receptor; thioesterase; inflammation;
KW multifactorial disease; autoimmune disease; infection;
KW nervous system disease; ss.
OS
XX Homo sapiens.
XX
PN WO200147944-A2.
XX
PD 05-JUL-2001.
XX
PF 28-DEC-2000; 2000WO-US35498.
XX
PR 28-DEC-1999; 99US-0173419.

```
PR 27-DEC-2000; 2000US-0173419.
XX (CURA-) CURAGEN CORP.
PA
XX Shimkets RA, Leach M;
XX WPI; 2001-465210/50.
XX
XX Polymorphic nucleic acids encoding e.g. amylases, cyclins, polymerases,
PT oncogenes and histones, useful for diagnosing and treating, e.g.
PT cancer, autoimmune diseases and infections -
XX
XX Claim 1; Page 2015; 4143pp; English.
XX
XX The present invention relates to oligonucleotides encoding polymorphic
CC variants of proteins related to amylases, amyloid proteins, angiotensin,
CC apoptosis related proteins, cadherin, cyclin, polymerase, oncogenes,
CC histones, kinases, colony stimulating factors, complement related
CC proteins, cytochromes, kinesins, cytokines, interferons, interleukins,
CC G-protein coupled receptors and thioesterases. The present sequence is
CC one such oligonucleotide. The oligonucleotides and the peptides encoded
CC by them may be used in the prevention, diagnosis and treatment of
CC diseases associated with inappropriate expression of the proteins listed
CC above. Disorders that may be prevented, diagnosed and/or treated include
CC multifactorial diseases with a genetic component, such as autoimmune
CC diseases (e.g. rheumatoid arthritis, multiple sclerosis, diabetes,
CC systemic lupus erythematosus and Grave's disease), inflammation, cancer
CC (e.g. cancers of the bladder, brain, breast, colon and kidney,
CC leukaemia), diseases of the nervous system and an infection of pathogenic
CC organisms.
XX
XX Sequence 50 BP; 8 A; 11 C; 11 G; 20 T; 0 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAL29004 ..
Align seg 1/1 to: AAL29004 from: 1 to: 50

488 GlyValSerGlnSerLeuSerGly 495
|||||
22 GGAGTTTCTCAGCGCTCTCGGT 45

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: AAD07056

seq_documentation_block:
ID AAD07056 standard; DNA; 310 BP.
XX
AC AAD07056;
XX
DT 06-AUG-2001 (first entry)
XX
DE Human DNA #1 related to the invention.
XX
KW Human; gene therapy; angiogenesis; vascularisation; cardiac hypertrophy;
KW congestive heart failure; CHF; myocardial infarction; aortic stenosis;
KW valvular regurgitation; vascular trauma; wound; burn; cancer; vaccine;
KW endothelial disease; ds.
XX
OS Homo sapiens.
XX
PN WO200125433-A2.
XX
PD 12-APR-2001.
XX
PF 05-OCT-2000; 2000WO-US27512.
XX
PR 07-OCT-1999; 99US-0158587.
```

```
PR 28-OCT-1999; 99US-0162611.
XX (GETH ) GENENTECH INC.
PA (CURA-) CURAGEN CORP.
XX
XX Gerritsen ME, Goddard A, Grimaldi JC, Mehraban F;
XX WPI; 2001-367229/38.
XX
XX Polypeptides critical for angiogenesis and vascularization, and the
PT nucleic acids encoding them, useful for treating conditions related to
PT inappropriate vascularization and angiogenesis -
XX
XX Disclosure; Page 187; 189pp; English.
XX
XX The patent discloses novel polypeptides critical for angiogenesis and
CC vascularisation and nucleic acids encoding them. The sequences of the
CC present invention are used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate angiogenesis and/or neo- or
CC cardio-vascularisation. They are used in the diagnosis, prevention
CC or treatment of angiogenic disorders such as cardiac hypertrophy,
CC congestive heart failure (CHF), myocardial infarction, aortic stenosis,
CC valvular regurgitation and vascular trauma such as wounds, burns or
CC surgery. They are also used in the treatment of cancer and endothelial
CC diseases. The proteins of the invention are used as antigens in the
CC production of antibodies against the angiogenesis and vascularisation
CC proteins and in assays to identify modulators (agonists and antagonists)
CC of their expression and activity. The antibodies and antagonists are
CC used to down regulate expression and activity of the protein of the
CC invention. The antibodies are also used as diagnostic agents for
CC detecting the presence of the proteins of the invention in samples.
CC The sequences of the invention are also used as vaccines and in gene
CC therapy. The present sequence is human DNA related to the invention.
XX
SQ Sequence 310 BP; 83 A; 83 C; 63 G; 80 T; 1 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAD07056 ..
Align seg 1/1 to: AAD07056 from: 1 to: 310

488 GlyValSerGlnSerLeuSerGly 495
|||||
158 GGAGTATCTCAAAGCCTTTCAGGC 181

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: AAK60152

seq_documentation_block:
ID AAK60152 standard; cDNA; 388 BP.
XX
AC AAK60152;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:5212.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
XX 09-AUG-2001.
XX
PD 17-JAN-2001; 2001WO-US01354.
XX
PR
```

PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI: 2001-483426/52.
P-PSDB; AAM87371.
Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and


```
PT metastasis -
XX Claim 1; SEQ ID NO 5212; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX Sequence 388 BP; 72 A; 94 C; 119 G; 100 T; 3 other;
SQ

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAK60152/rev ..

Align seg 1/1 to reverse of: AAK60152 from: 1 to: 388

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
361 GAAGGGTTACAGGAGGTTTAAAG 338

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAK56055

seq_documentation_block:
ID AAK56055 standard; DNA; 413 BP.
XX
XX AAK56055;
AC
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #186.
XX
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN W0200053724-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06112.
XX
XX 11-MAR-1999; 99US-0266513.
XX 18-AUG-1999; 99US-0149485.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX WPI; 2000-579369/54.
XX

New isolated polynucleotide encoding a plant transcription factor for
producing a plant e.g. a woody plant, preferably eucalyptus or pine,
having modified gene expression or modified activity of a polypeptide
-
Claim 1; Page 93; 747pp; English.
XX
XX The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
XX Sequence 413 BP; 99 A; 92 C; 120 G; 102 T; 0 other;
SQ

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAC56055/rev ..

Align seg 1/1 to reverse of: AAC56055 from: 1 to: 413

429 SerGlySerSerThrIleAlaAla 436
|||||
311 TCAGGTAGTTCAACCATCGCAGCC 288

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAK03891

seq_documentation_block:
ID AAC03891 standard; cDNA; 447 BP.
XX
XX AAC03891;
AC
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 3889.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX P-PSDB; AAG03885.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
```

PS Claim 1; SEQ ID 3889; 71pp + CD-ROM; English.

XX The present sequence is one of a large number of 5' ESTs derived from

CC mRNAs encoding secreted proteins. An ORF has been identified within the

CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond

CC mainly to the 3' untranslated region (UTR) of the mRNA because they are

CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not

CC well suited for isolating cDNA sequences derived from the 5' ends of

CC mRNAs and even in those cases where longer cDNA sequences have been

CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from

CC mRNAs with intact 5' ends and can therefore be used to obtain full length

CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,

CC gene therapy and chromosome mapping procedures. They are used to obtain

CC upstream regulatory sequences and to design expression and secretion

XX vectors.

SQ Sequence 447 BP; 111 A; 130 C; 119 G; 84 T; 3 other;

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAC03891 ..

Align seg 1/1 to: AAC03891 from: 1 to: 447

113 ProlysGluGlyLeuGlnGluGly 120

|||||

78 CCCAAGAAGGGCTCCCAAGAGGC 101

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:AAQ79427

seq_documentation_block:

ID_AAQ79427 standard; cDNA; 450 BP.

XX AC AQAQ79427;

XX 03-JUN-1995 (first entry)

XX Human islet inducible nitric oxide synthase (iNOS) cDNA 3' end.

XX Nitric oxide synthase; diabetes mellitus; islet; ss.

XX Homo sapiens.

XX WO9424269-A.

XX 27-OCT-1994.

XX 11-APR-1994; 94WO-DR00146.

XX 16-APR-1993; 93DK-0000433.

XX (NOVO) NOVO-NORDISK AS.

XX Karlsen AE;

XX WPI; 1994-341851/42.

XX New DNA encoding inducible pancreatic islet nitric oxide synthase

PT - and related vectors and transformed cells, useful for

PT identifying specific inhibitors for treatment or prevention of

PT insulin dependent diabetes mellitus

XX Disclosure; Page 26-27; 36pp; English.

XX Human islets were incubated in a mixture of IL-1, TNF-alpha

CC and IFN-gamma. The human islet iNOS was cloned by RT-PCR on

CC the isolated mRNA with primers based on the human

CC hepatocyte sequence. The sequence of the primers is given in

CC AAQ79424 and AAQ79425. The clone human islet iNOS was sequenced.

CC The 5' UTR and translated human islet iNOS is given in AAQ79426,

CC and the 3' end in AAQ79427.

SQ Sequence 450 BP; 89 A; 142 C; 116 G; 103 T; 0 other;

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAQ79427/rev ..

Align seg 1/1 to reverse of: AAQ79427 from: 1 to: 450

115 GluGlyLeuGlnGluGlyLeuLys 122

|||||

146 GAGGCGCTACAGGAGGGCTTAAG 123

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:ABA62517

seq_documentation_block:

ID_ABA62517 standard; DNA; 511 BP.

XX AC ABA62517;

XX 01-FEB-2002 (first entry)

XX Human foetal liver single exon nucleic acid probe #10822.

DE Human; foetal liver; gene expression; single exon nucleic acid probe; ss.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human fetal liver -

XX Claim 1; SEQ ID NO 10822; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC fetal liver. The present sequence is a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x ABA62517/rev ..

Align seg 1/1 to reverse of: ABA62517 from: 1 to: 511

47 GluSerLeuGlnSerValLeuPro 54
|||||
462 GAGAGCCTCCAGTCTGTCCTTCT 439

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:ABA29847

seq_documentation_block:
ID ABA29847 standard; DNA; 511 BP.
XX
AC ABA29847;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #8313 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.

OS Homo sapiens.

XX

XX WO200157274-A2.

XX

XX 09-AUG-2001.

XX

XX 30-JAN-2001; 2001WO-US00666.

XX

XX 04-FEB-2000; 2000US-0180312.

XX

XX 26-MAY-2000; 2000US-0207456.

XX

XX 30-JUN-2000; 2000US-0608408.

XX

XX 03-AUG-2000; 2000US-0632366.

XX

XX 21-SEP-2000; 2000US-0234687.

XX

XX 27-SEP-2000; 2000US-0236359.

XX

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA

XX

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human hearts -

XX Claim 1; SEQ ID No 8313; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.

XX Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x ABA29847/rev ..

Align seg 1/1 to reverse of: ABA29847 from: 1 to: 511

47 GluSerLeuGlnSerValLeuPro 54
|||||
462 GAGAGCCTCCAGTCTGTCCTTCT 439

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAK10872

seq_documentation_block:
ID AAK10872 standard; DNA; 511 BP.

XX

XX AAK10872;

XX

XX 05-NOV-2001 (first entry)

XX

XX Human brain expressed single exon probe SEQ ID NO: 10863.

XX

XX Human; brain expressed exon; gene expression analysis; probe;

XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX epilepsy; cancer; ss.

XX

XX Homo sapiens.

XX

XX WO200157275-A2.

XX

XX 09-AUG-2001.

XX

XX 30-JAN-2001; 2001WO-US00667.

XX

XX 04-FEB-2000; 2000US-0180312.

XX

XX 26-MAY-2000; 2000US-0207456.

XX

XX 30-JUN-2000; 2000US-0608408.

XX

XX 03-AUG-2000; 2000US-0632366.

XX

XX 21-SEP-2000; 2000US-0234687.

XX

XX 27-SEP-2000; 2000US-0236359.

XX

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human brains -

XX Example 4; SEQ ID NO: 10863; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention.

XX Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
US-09-701-711-2 x AAK10872/rev ..
Align seg 1/1 to reverse of: AAK10872 from: 1 to: 511
47 GluSerLeuGlnSerValleuPro 54
|||||
462 GAGAGCCTCCAGTCGTCTCCT 439

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AAK36734
seq_documentation_block:
ID AAK36734 standard; DNA; 511 BP.
XX
AC AAK36734;
XX
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 11291.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow.
XX
PS Example 4; SEQ ID NO: 11291; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAK36734/rev ..
Align seg 1/1 to reverse of: AAK36734 from: 1 to: 511
47 GluSerLeuGlnSerValleuPro 54
|||||
462 GAGAGCCTCCAGTCGTCTCCT 439
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AAI17584
seq_documentation_block:
ID AAI17584 standard; DNA; 511 BP.
XX
AC AAI17584;
XX
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #7517 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells.
XX
PS Claim 25; SEQ ID NO 7517; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENPs). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAI17584/rev ..
Align seg 1/1 to reverse of: AAI17584 from: 1 to: 511
47 GluSerLeuGlnSerValleuPro 54
|||||
462 GAGAGCCTCCAGTCGTCTCCT 439

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AAI42500
seq_documentation_block:
```

ID AAI42500 standard; DNA; 511 BP.
XX
AC AAI42500;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #1186 used to measure gene expression in human placenta sample.
XX
KW Probe: microarray; human; placenta; antenatal diagnosis;
KW genetic disorder; ss.
XX Homo sapiens.
XX
PN W0200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 11186; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 511 BP; 114 A; 129 C; 121 G; 147 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAI42500/rev ..

Align seg 1/1 to reverse of: AAI42500 from: 1 to: 511

47 GluSerLeuGlnSerValLeuPro 54
|||||
462 GAGAGCCTCCAGTCGTGCTCCT 439

seq_name: /SDSL/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AAS42998

seq_documentation_block:
ID AAS42998 standard; cDNA; 583 BP.
XX
AC AAS42998;
XX
DT 18-DEC-2001 (first entry)
XX
DE DNA encoding G protein-coupled receptor, nGPCR-2102.
XX
KW Human; mental disorder; thyroid disease; renal failure; anorexia;

KW inflammatory condition; Crohn's disease; rheumatoid arthritis; HIV;
KW autoimmune disorder; schizophrenia; migraine; stroke; dementia; obesity;
KW depression; Parkinson's disease; Alzheimer's disease; viral infection;
KW Huntington's disease; human immunodeficiency virus; type 2 diabetes;
KW anorexia; hypotension; hypertension; thrombosis; myocardial infarction;
KW atherosclerosis; cancer; sexual dysfunction; G protein-coupled receptor;
KW nGPCR; ss.
XX
OS Homo sapiens.
XX
PN W0200162924-A2.
XX
PD 30-AUG-2001.
XX
PF 23-FEB-2001; 2001WO-US05989.
XX
XX 24-FEB-2000; 2000US-0184602.
PR 24-FEB-2000; 2000US-0184604.
PR 24-FEB-2000; 2000US-0184606.
PR 24-FEB-2000; 2000US-0184689.
PR 24-FEB-2000; 2000US-0184690.
PR 24-FEB-2000; 2000US-0184710.
PR 24-FEB-2000; 2000US-0184712.
PR 24-FEB-2000; 2000US-0184715.
PR 24-FEB-2000; 2000US-0184716.
PR 24-FEB-2000; 2000US-0184725.
PR 24-FEB-2000; 2000US-0184822.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Vogeli G, Wood LS, Parodi LA, Lind P;
XX
XX WPI; 2001-570632/64.
DR P-PSDB; AAU25688.
XX
PT Novel nucleic acid and encoded nGPCR-x, used to screen for compounds
PT for use in the treatment of mental disorders, such as Alzheimer's
XX disease, or Parkinson's disease -
XX
PS Claim 3; Page 81; 263pp; English.
XX
CC The invention relates to novel isolated human G protein-coupled
CC receptors (nGPCR-x). The nGPCR-x can be used for screening compounds
CC which can be used to treat mental disorders, thyroid disease, renal
CC failure, inflammatory conditions such as Crohn's disease, rheumatoid
CC arthritis, autoimmune disorders, schizophrenia, migraine, stroke,
CC dementia, depression, Parkinson's disease, Alzheimer's disease, and
CC Huntington's disease. They may also be used for treating viral infections
CC such as human immunodeficiency virus (HIV), type 2 diabetes, obesity,
CC anorexia, hypotension, hypertension, thrombosis, myocardial infarction,
CC atherosclerosis, cancer, and sexual dysfunction. AAS42927-AAS43036
CC represent the coding sequences of novel human G protein-coupled
CC receptors, nGPCR-2031 to nGPCR-2140 respectively, as described in the
CC invention.
XX
SQ Sequence 583 BP; 190 A; 86 C; 102 G; 205 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAS42998 ..

Align seg 1/1 to: AAS42998 from: 1 to: 583

310 ThrThrAsnAsnIleSerThrIys 317
|||||
22 ACAACAAACAATATTTCTACTAAA 45

seq_name: /SDSL/gcgdata/geneseq/geneseq-embl/NA1998.DAT:AAV35370

seq_documentation_block:
ID AAV35370 standard; DNA; 604 BP.
AC AAV35370;
XX
XX 09-OCT-1998 (first entry)
XX
XX Human INOS genomic DNA fragment #1.
XX
XX Inducible nitric oxide synthase; hINOS; human; regulator; expression;
KW treatment; inflammation; sepsis; antitumour agent; inhibitor; ds.
XX
XX Homo sapiens.
XX
XX WO9812313-A1.
XX
XX 26-MAR-1998.
XX
XX 18-SEP-1997; 97WO-JP03303.
XX
XX 20-SEP-1996; 96JP-0250697.
XX
XX (SUNR) SUNTORY LTD.
XX
XX Nunokawa Y, Oikawa S, Tanaka S;
XX
XX WPI; 1998-230314/20.
XX
XX Screening for regulators of human inducible nitric oxide synthase
PT expression - using human cell line transformed with a reporter gene
PT flanked by 5'-promoter and 3'-non-translated regions of the hINOS
PT gene
XX
XX Disclosure; Fig 2; 56pp; Japanese.
XX
XX This sequence is a fragment of human inducible nitric oxide synthase
CC (hINOS) which can be used to screen for potential hINOS expression
CC regulators. Such regulators have an effect on the modification of the
CC expression (in the presence of a cytokine inducer) of a reporter gene in
CC a human cell line transformed with an expression vector containing (in
CC order from 5'-end): the 5'-flanking region of the hINOS gene containing
CC 544 bases, the reporter gene and the 3'-untranslated region of the hINOS
CC gene. Suitable reporter genes include chloramphenicol acetyltransferase
CC (CAT), beta-galactosidase (beta -gal) and luciferase. Compounds which
CC influence the expression of the hINOS gene are useful for the treatment
CC of inflammation and sepsis, as antitumour agents and for the inhibition
CC of re-angio-stenosis, and other conditions involving abnormalities of
CC hINOS expression.
XX
SQ Sequence 604 BP; 158 A; 150 C; 146 G; 150 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAV35370 ..

Align seg 1/1 to: AAV35370 from: 1 to: 604

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
107 GAGGGCCTACAGGAGGGGTTAAAG 130

seq_name: /SIDS1/gcdata/geneseq/geneseq-embl/NA1999.DAT:AAZ32098

seq_documentation_block:

ID AAZ32098 standard; DNA; 604 BP.

XX

AC AAZ32098;

XX 11-JAN-2000 (first entry)
DT
XX
XX HINOS exon 26 nucleotide sequence.
XX
XX Human induced nitrogen monoxide synthase; hINOS; untranslated region;
KW NF-kappaB; expression; activation; regulation; ds.
XX
XX Homo sapiens.
XX
XX JP11266872-A.
XX
XX 05-OCT-1999.
XX
XX 20-MAR-1998; 98JP-0090664.
XX
XX 20-MAR-1998; 98JP-0090664.
XX
XX (SUNR) SUNTORY LTD.
XX
XX WPI; 1999-613779/53.
XX
XX Method of screening a substance controlling activation of NF-kappaB -
PT having increased sensitivity
PT
XX
XX Disclosure; Fig 2; 16pp; Japanese.
XX
XX The present invention describes an expression-controlling sequence
CC containing a NF-kappaB recognising sequence, at least part of
CC 3'-untranslated region (3'-UTR) of human-induced nitrogen monoxide
CC synthase (hINOS) and at least part of 3'-flanking region. Also described
CC are: (a) an expression vector containing the above expression-controlling
CC sequence; (b) an expression vector containing (1) 5'-flanking region of
CC hINOS gene containing a promoter region, (2) a reporter gene and (3) the
CC above expression-controlling sequence in this order from the 5'-side;
CC (c) a cell transformed by the above expression vector; (d) a method for
CC screening a substance controlling activation of NF-kappaB in which a cell
CC which has the above expression-controlling sequence and can detect the
CC activation of NF-kappaB is treated with a sample to observe the change in
CC the expressed amount of the reporter gene; (e) a kit for screening a
CC substance controlling activation of NF-kappaB containing the above cell;
CC (f) a compound controlling activation of NF-kappaB prepared by using the
CC above screening method; and (g) a drug composition for treating diseases
CC caused by activation of NF-kappaB containing, with the above compound as
CC the active component. The method can evaluate a compound controlling
CC activation of NF-kappaB easily in a high sensitivity. The present
CC sequence is used in the exemplification of the present invention.
XX
SQ Sequence 604 BP; 158 A; 150 C; 146 G; 150 T; 0 other;

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAZ32098 ..

Align seg 1/1 to: AAZ32098 from: 1 to: 604

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
107 GAGGGCCTACAGGAGGGGTTAAAG 130

seq_name: /SIDS1/gcdata/geneseq/geneseq-embl/NA2000.DAT:AAFI6109

seq_documentation_block:

ID AAFI6109 standard; cDNA; 842 BP.

XX

AC AAFI6109;

XX

DT 13-MAR-2001 (first entry)

XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:544.
DE
XX
XX Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200055174-A1.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05988.
PF
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX (ROSE/) ROSEN C A.
PI
XX
XX Rosen CA, Ruben SM;
XX
XX WPI; 2000-587513/55.
DR
XX
XX P-PSDB; AAB56906.
DR
XX
XX Prostate cancer associated gene sequences, referred to as prostate
PT cancer antigens, useful for treatment, prevention, and diagnosis of
PT disorders such as prostate cancer -
XX
XX
PS Claim 1; Page 1004; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC The prostate cancer antigens can have neuroprotective, cytostatic,
CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC nephrotropic, antineoplastic, gynaecological and antibacterial activities,
CC and can be used in gene therapy. The prostate cancer antigen
CC polynucleotides may be used for detection of prostate cancer
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The prostate cancer antigens may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC AAB57303 represent sequences used in the exemplification of the present
CC invention.
XX
XX
SQ Sequence 842 BP; 257 A; 201 C; 205 G; 179 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAF16109 ..

Align seg 1/1 to: AAF16109 from: 1 to: 842

113 ProLysGluGlyLeuGlnGluGly 120

|||||

101 CCCAAGAGGGCTCCAGAGGC 124

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT: AAS81839

seq_documentation_block:

ID AAS81839 standard; cDNA; 863 BP.

XX AAS81839;

AC
XX

DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #17643.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX
XX 23-AUG-2000; 2000US-0649167.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR
XX
XX P-PSDB; ABG17652.
DR
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX
PS Claim 1; SEQ ID No 17643; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 863 BP; 222 A; 214 C; 231 G; 196 T; 0 other;

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AAS81839 ..

Align seg 1/1 to: AAS81839 from: 1 to: 863

370 ValLeuArgArgGluMetArgGln 377

|||||

6 GTCTTCGCTCGGAATGCGTCAG 29

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT: AAV35383

seq_documentation_block:
ID AAV35383 standard; DNA; 1026 BP.
AC AAV35383;
XX
XX
DT 09-OCT-1998 (first entry)
XX
DE Human INOS genomic DNA fragment 3'-end #1.
XX
KW Inducible nitric oxide synthase; hINOS; human; regulator; expression;
KW treatment; inflammation; sepsis; antitumour agent; inhibitor; ds.
XX
OS Homo sapiens.
XX
PN WO9812313-A1.
XX
XX 26-MAR-1998.
PD
XX 18-SEP-1997; 97WO-JP03303.
PF
XX 20-SEP-1996; 96JP-0250697.
PR
XX (SUNR) SUNTORY LTD.
PA
XX Nunokawa Y, Oikawa S, Tanaka S;
PI
XX WPI; 1998-230314/20.
XX
XX Screening for regulators of human inducible nitric oxide synthase
PT expression - using human cell line transformed with a reporter gene
PT flanked by 5'-promoter and 3'-non-translated regions of the hINOS
PT gene
XX
XX Disclosure; Page 28-29; 56pp; Japanese.
PS
XX This sequence is a 3'-end fragment of human inducible nitric oxide
CC synthase (hINOS) which can be used to screen for potential hINOS
CC expression regulators. Such regulators have an effect on the modification
CC of the expression (in the presence of a cytokine inducer) of a reporter
CC gene in a human cell line transformed with an expression vector
CC containing (in order from 5'-end): the 5'-flanking region of the hINOS
CC gene containing 544 bases, the reporter gene and the 3'-untranslated
CC region of the hINOS gene. Suitable reporter genes include chloramphenicol
CC acetyltransferase (CAT), beta-galactosidase (beta -Gal) and luciferase.
CC Compounds which influence the expression of the hINOS gene are useful for
CC the treatment of inflammation and sepsis, as antitumour agents and for
CC the inhibition of re-angio-stenosis, and other conditions involving
CC abnormalities of hINOS expression.
XX
SQ Sequence 1026 BP; 219 A; 259 C; 299 G; 249 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAV35383 ..
Align seg 1/1 to: AAV35383 from: 1 to: 1026

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
2 GAGGGCCTACAGGAGGGGTTAAAG 25

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAV32093
seq_documentation_block:
ID AAV32093 standard; DNA; 1026 BP.
XX
AC AAV32093;
XX

DT 11-JAN-2000 (first entry)
XX
DE hINOS 3'-untranslated region SEQ ID NO:13.
XX
KW Human induced nitrogen monoxide synthase; hINOS; untranslated region;
KW NF-kappaB; expression; activation; regulation; ds.
XX
OS Homo sapiens.
XX
PN JP11266872-A.
XX
XX 05-OCT-1999.
PD
XX 20-MAR-1998; 98JP-0090664.
PF
XX 20-MAR-1998; 98JP-0090664.
PR
XX (SUNR) SUNTORY LTD.
PA
XX WPI; 1999-613779/53.
DR
XX Method of screening a substance controlling activation of NF-kappaB -
PT having increased sensitivity
PT
XX Example 1; Page 10-11; 16pp; Japanese.
PS
XX The present invention describes an expression-controlling sequence
CC containing a NF-kappaB recognising sequence, at least part of
CC 3'-untranslated region (3'-UTR) of human-induced nitrogen monoxide
CC synthase (hINOS) and at least part of 3'-flanking region. Also described
CC are: (a) an expression vector containing the above expression-controlling
CC sequence; (b) an expression vector containing (1) 5'-flanking region of
CC hINOS gene containing a promoter region, (2) a reporter gene and (3) the
CC above expression-controlling sequence in this order from the 5'-side;
CC (c) a cell transformed by the above expression vector; (d) a method for
CC screening a substance controlling activation of NF-kappaB in which a cell
CC which has the above expression-controlling sequence and can detect the
CC activation of NF-kappaB is treated with a sample to observe the change in
CC the expressed amount of the reporter gene; (e) a kit for screening a
CC substance controlling activation of NF-kappaB containing the above cell;
CC (f) a compound controlling activation of NF-kappaB prepared by using the
CC above screening method; and (g) a drug composition for treating diseases
CC caused by activation of NF-kappaB containing, with the above compound as
CC the active component. The method can evaluate a compound controlling
CC activation of NF-kappaB easily in a high sensitivity. The present
CC sequence represents hINOS 3'-untranslated region from the present
CC invention.
XX
SQ Sequence 1026 BP; 219 A; 259 C; 299 G; 249 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAV32093 ..
Align seg 1/1 to: AAV32093 from: 1 to: 1026

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
2 GAGGGCCTACAGGAGGGGTTAAAG 25

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAC38511
seq_documentation_block:
ID AAC38511 standard; DNA; 1143 BP.
XX
AC AAC38511;
XX
DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 21235.
XX DE
XX XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX OS
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX PD
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PF 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 29-MAR-1999; 99US-0126264.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.
XX PR 08-APR-1999; 99US-0128714.
XX PR 16-APR-1999; 99US-0129845.
XX PR 19-APR-1999; 99US-0130077.
XX PR 21-APR-1999; 99US-0130449.
XX PR 23-APR-1999; 99US-0130510.
XX PR 23-APR-1999; 99US-0130891.
XX PR 28-APR-1999; 99US-0131449.
XX PR 30-APR-1999; 99US-0132048.
XX PR 30-APR-1999; 99US-0132407.
XX PR 04-MAY-1999; 99US-0132484.
XX PR 05-MAY-1999; 99US-0132485.
XX PR 06-MAY-1999; 99US-0132486.
XX PR 07-MAY-1999; 99US-0132487.
XX PR 11-MAY-1999; 99US-0132863.
XX PR 14-MAY-1999; 99US-0134256.
XX PR 14-MAY-1999; 99US-0134218.
XX PR 14-MAY-1999; 99US-0134219.
XX PR 14-MAY-1999; 99US-0134221.
XX PR 14-MAY-1999; 99US-0134370.
XX PR 18-MAY-1999; 99US-0134768.
XX PR 19-MAY-1999; 99US-0134941.
XX PR 20-MAY-1999; 99US-0135124.
XX PR 21-MAY-1999; 99US-0135353.
XX PR 24-MAY-1999; 99US-0135629.
XX PR 25-MAY-1999; 99US-0136021.
XX PR 27-MAY-1999; 99US-0136392.
XX PR 28-MAY-1999; 99US-0136782.
XX PR 01-JUN-1999; 99US-0137222.
XX PR 03-JUN-1999; 99US-0137528.
XX PR 04-JUN-1999; 99US-0137502.
XX PR 07-JUN-1999; 99US-0137724.
XX PR 08-JUN-1999; 99US-0138094.
XX PR 10-JUN-1999; 99US-0138540.
XX PR 10-JUN-1999; 99US-0138847.
XX PR 14-JUN-1999; 99US-0139119.
XX PR 16-JUN-1999; 99US-0139452.
XX PR 16-JUN-1999; 99US-0139453.
XX PR 17-JUN-1999; 99US-0139492.
XX PR 18-JUN-1999; 99US-0139454.
XX PR 18-JUN-1999; 99US-0139455.
XX PR 18-JUN-1999; 99US-0139456.
XX PR 18-JUN-1999; 99US-0139457.
XX PR 18-JUN-1999; 99US-0139458.
XX PR 18-JUN-1999; 99US-0139459.
XX PR 18-JUN-1999; 99US-0139460.
XX PR 18-JUN-1999; 99US-0139461.
XX PR 18-JUN-1999; 99US-0139462.
XX PR 18-JUN-1999; 99US-0139463.
XX PR 18-JUN-1999; 99US-0139750.
XX PR 18-JUN-1999; 99US-0139763.
XX PR 21-JUN-1999; 99US-0139817.
XX PR 22-JUN-1999; 99US-0139899.
XX PR 23-JUN-1999; 99US-0140353.
XX PR 23-JUN-1999; 99US-0140354.
XX PR 24-JUN-1999; 99US-0140695.
XX PR 28-JUN-1999; 99US-0140823.
XX PR 29-JUN-1999; 99US-0140991.
XX PR 30-JUN-1999; 99US-0141287.
XX PR 01-JUL-1999; 99US-0141842.
XX PR 02-JUL-1999; 99US-0142154.
XX PR 02-JUL-1999; 99US-0142055.
XX PR 06-JUL-1999; 99US-0142390.
XX PR 08-JUL-1999; 99US-0142803.
XX PR 09-JUL-1999; 99US-0142920.
XX PR 12-JUL-1999; 99US-0142977.
XX PR 13-JUL-1999; 99US-0143542.
XX PR 14-JUL-1999; 99US-0143624.
XX PR 15-JUL-1999; 99US-0144005.
XX PR 16-JUL-1999; 99US-0144085.
XX PR 16-JUL-1999; 99US-0144086.
XX PR 19-JUL-1999; 99US-0144325.
XX PR 19-JUL-1999; 99US-0144331.
XX PR 19-JUL-1999; 99US-0144332.
XX PR 19-JUL-1999; 99US-0144333.
XX PR 19-JUL-1999; 99US-0144334.
XX PR 19-JUL-1999; 99US-0144335.
XX PR 20-JUL-1999; 99US-0144352.
XX PR 20-JUL-1999; 99US-0144632.
XX PR 20-JUL-1999; 99US-0144884.
XX PR 21-JUL-1999; 99US-0144814.
XX PR 21-JUL-1999; 99US-0145086.
XX PR 22-JUL-1999; 99US-0145085.
XX PR 22-JUL-1999; 99US-0145087.
XX PR 22-JUL-1999; 99US-0145089.
XX PR 22-JUL-1999; 99US-0145192.
XX PR 23-JUL-1999; 99US-0145145.
XX PR 23-JUL-1999; 99US-0145218.
XX PR 23-JUL-1999; 99US-0145224.
XX PR 26-JUL-1999; 99US-0145276.
XX PR 27-JUL-1999; 99US-0145913.
XX PR 27-JUL-1999; 99US-0145918.
XX PR 27-JUL-1999; 99US-0145951.
XX PR 28-JUL-1999; 99US-0145951.
XX PR 02-AUG-1999; 99US-0146386.
XX PR 02-AUG-1999; 99US-0146388.
XX PR 02-AUG-1999; 99US-0146389.
XX PR 03-AUG-1999; 99US-0147038.
XX PR 04-AUG-1999; 99US-0147204.
XX PR 04-AUG-1999; 99US-0147302.
XX PR 05-AUG-1999; 99US-0147192.
XX PR 05-AUG-1999; 99US-0147260.
XX PR 06-AUG-1999; 99US-0147303.
XX PR 06-AUG-1999; 99US-0147416.
XX PR 09-AUG-1999; 99US-0147493.
XX PR 09-AUG-1999; 99US-0147935.
XX PR 10-AUG-1999; 99US-0148171.
XX PR 11-AUG-1999; 99US-0148319.
XX PR 12-AUG-1999; 99US-0148341.
XX PR 13-AUG-1999; 99US-0148565.
XX PR 13-AUG-1999; 99US-0148684.
XX PR 16-AUG-1999; 99US-0149368.
XX PR 17-AUG-1999; 99US-0149175.
XX PR 18-AUG-1999; 99US-0149426.
XX PR 20-AUG-1999; 99US-0149722.
XX PR 20-AUG-1999; 99US-0149723.
XX PR 23-AUG-1999; 99US-0149929.
XX PR 23-AUG-1999; 99US-0149902.
XX PR 23-AUG-1999; 99US-0149930.
XX PR 25-AUG-1999; 99US-0150566.
XX PR 26-AUG-1999; 99US-0150884.
XX PR 27-AUG-1999; 99US-0151065.
XX PR 27-AUG-1999; 99US-0151066.

```
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 08-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAC38511/rev ..

Align seg 1/1 to reverse of: AAC38511 from: 1 to: 1143

666 GlnSerLeuGlyProArgSer 673
|||||
467 CAYCCTCTCGGTCGCAAGGAGT 444

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:AAQ38956
seq_documentation_block:
ID AAQ38956 standard; DNA; 1637 BP.
XX AC AAQ38956;
XX DT 28-JUL-1993 (first entry)
XX
```

```
DE Lipase gene #2.
XX
KW Lipase; Pseudomonas; mutant; stability; specificity; ss.
XX
OS Pseudomonas nov. sp. No.109 FERM 3025.
XX
FH Key Location/Qualifiers
FT CDS 13..1380
FT /*tag= a
XX
PN JP05056786-A.
XX
XX 09-MAR-1993.
XX
XX 26-AUG-1991; 91JP-0296730.
XX
XX 26-AUG-1991; 91JP-0296730.
XX
XX (NAGA-) NAGASE SETKAGAKU KOGYO KK.
XX
XX WPI; 1993-120382/15.
XX
XX P-PSDB; AAR35205.
XX
XX DNA fragment contg. lipase producing gene from Pseudomonas - used
XX to produce large amt. of mutated lipase with increased stability
XX
XX Claim 1; Page 8-11; 13pp; Japanese.
XX
XX The sequences given in AAQ38955-56 encode lipase producing genes from
XX a Pseudomonas species, pref. P. nov. sp. No. 109 FERM 3025. These
XX genes can be used to produce large amount of lipase. They can also be
XX used to produce a mutated lipase which has increased stability or a
XX change in specificity.
XX
XX Sequence 1637 BP; 276 A; 575 C; 539 G; 247 T; 0 other;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAQ38956 ..

Align seg 1/1 to: AAQ38956 from: 1 to: 1637

376 ArgGlnLeuGlyAlaLeuAla 383
|||||
1531 CGACAGCTTGAGCGCGCTGCG 1554

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:AAH74178
seq_documentation_block:
ID AAH74178 standard; cDNA; 1928 BP.
XX
XX AC AAH74178;
XX
XX 15-OCT-2001 (first entry)
XX
XX Rho family active site of GTPase coding sequence.
DE
XX Rho family active site; GTPase; gene therapy; cancer; haemopathy;
KW HIV infection; immunological disease; inflammation; ss.
XX
XX Unidentified.
XX
XX WO200155420-A1.
XX
XX 02-AUG-2001.
XX
XX 15-JAN-2001; 2001WO-CN00047.
XX
```

PR 26-JAN-2000; 2000CN-01111502.
PA (BIOD-) BIODOR GENE TECHNOLOGY LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WPI; 2001-483260/52.
DR P-PSDB; AAG65047.
XX
XX New Rho family active site of GTPase for diagnosing and treating
PT malignant tumor, hemopathy, human immunodeficiency virus infection,
PT immunological diseases and various inflammations -
XX
PS Claim 6; Page 29-30; 37pp; Chinese.
XX
XX The present invention provides the protein and coding sequence of the Rho
CC family active site of GTPase. The sequences can be used in the treatment
CC of cancer, haemopathy, HIV infection, immunological diseases and
CC inflammation. The present sequence is the coding sequence of the
CC invention.
XX
XX Sequence 1928 BP; 660 A; 370 C; 404 G; 494 T; 0 other;
SQ
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x AAH74178 ..
Align seg 1/1 to: AAH74178 from: 1 to: 1928
488 GlyValSerGlnSerLeuSerGly 495
821 GGAGTATCTCAAGCGCTTCAGGC 844
seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2000.DAT: AAC78114
seq_documentation_block:
ID AAC78114 standard; cDNA; 1930 BP.
XX
AC AAC78114;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human cancer associated gene sequence SEQ ID NO:508.
XX
KW Human; cancer associated gene; cancer antigen; detection; cancer;
KW diagnosis; cytostatic; proliferative; vulnerable; immunomodulator;
KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; neutropic;
KW vasotropic; antipsoriatic; antidiabetic; antidiabetic; neutropic;
KW immune disorder; haematopoietic cell disorder; gene therapy; inflammation;
KW allergic reaction; graft versus host disease; autoimmune disorder;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
OS Homo sapiens.
XX
XX WO200055350-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US05882.
XX
XX 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
PI

XX WPI; 2000-587533/55.
DR P-PSDB; AAB43905.
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
XX Claim 1; Page 1037-1038; 2352pp; English.
XX
XX AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerable; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC neutropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 1930 BP; 634 A; 291 C; 320 G; 682 T; 3 other;
alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x AAC78114/rev ..
Align seg 1/1 to reverse of: AAC78114 from: 1 to: 1930
703 LeuLeuLeuProLeuProPhelys 710
|||||
1719 CTTATCTTCCACTTCCCTTTAAA 1696
seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001B.DAT: ABL05175
seq_documentation_block:
ID ABL05175 standard; cDNA; 2009 BP.
XX
AC ABL05175;
XX
DT 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 10007.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW Drosophila; gene; ss.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX

PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB61072.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 10007; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2009 BP; 647 A; 459 C; 414 G; 489 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x ABL05175/rev ..

Align seg 1/1 to reverse of: ABL05175 from: 1 to: 2009

525 IleAspGluAsnGlnArgIleSer 532
|||||
64 ATCGATGAAACCAACGAAATTCG 41

seq_name: /SIDSl/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT: AAS81841

seq_documentation_block:
ID AAS81841 standard; cDNA; 2057 BP.
XX
AC AAS81841;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #17645.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR P-PSDB; ABB17654.
DR

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID NO 17645; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2057 BP; 505 A; 495 C; 550 G; 507 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAS81841 ..

Align seg 1/1 to: AAS81841 from: 1 to: 2057

370 ValLeuArgArgGluMetArgGln 377
|||||
1231 GTCTCGCTCGCGAAATGCGTCAG 1254

seq_name: /SIDSl/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT: ABL24790

seq_documentation_block:
ID ABL24790 standard; DNA; 2341 BP.
XX
AC ABL24790;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 25843.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA

XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
DR
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
PS Claim 1; SEQ ID NO 25843; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2341 BP; 616 A; 449 C; 549 G; 727 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x ABL24790/rev ..

Align seg 1/1 to reverse of: ABL24790 from: 1 to: 2341

696 LeuAlaThrPheGlySerGluLeu 703
|||||
654 TTAGCAACATTTGGCAGCGAGCTA 631

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: AAD07053

seq_documentation_block:
ID AAD07053 standard; cDNA; 2366 BP.
XX
AC AAD07053;
XX
DT 06-AUG-2001 (first entry)
DE
XX Human DNA-C-MG-72-1776 cDNA clone encoding PRO-C-MG.72 protein.
XX
XX Human; gene therapy; angiogenesis; vascularisation; cardiac hypertrophy;
KW congestive heart failure; CHF; myocardial infarction; aortic stenosis;
KW valvular regurgitation; vascular trauma; wound; burn; cancer; vaccine;
KW endothelial disease; DNA-C-MG.72-1776 cDNA; PRO-C-MG.72 protein; ss.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 71..2062
FT /*tag= a
FT /product= "Human PRO-C-MG.72 protein"
XX
XX WO200125433-A2.
XX
XX
PD 12-APR-2001.
XX
XX 05-OCT-2000; 2000WO-US27512.
XX
XX 07-OCT-1999; 99US-0158587.
XX
XX 28-OCT-1999; 99US-0162611.
XX
XX (GETH) GENENTECH INC.

PA (CURA-) CURAGEN CORP.
XX
XX Gerritsen ME, Goddard A, Grimaldi JC, Mehraban F;
PI
XX
DR WPI; 2001-367229/38.
DR P-PSDB; AAE02777.
XX
XX Polypeptides critical for angiogenesis and vascularization, and the
PT nucleic acids encoding them, useful for treating conditions related to
PT inappropriate vascularization and angiogenesis -
XX
PS Claim 10; Page 172-175; 189pp; English.
XX
CC The patent discloses novel polypeptides critical for angiogenesis and
CC vascularisation and nucleic acids encoding them. The sequences of the
CC present invention are used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate angiogenesis and/or neo- or
CC cardio-vascularisation. They are used in the diagnosis, prevention
CC or treatment of angio-genic disorders such as cardiac hypertrophy,
CC congestive heart failure (CHF), myocardial infarction, aortic stenosis,
CC valvular regurgitation and vascular trauma such as wounds, burns or
CC surgery. They are also used in the treatment of cancer and endothelial
CC diseases. The proteins of the invention are used as antigens in the
CC production of antibodies against the angiogenesis and vascularisation
CC proteins and in assays to identify modulators (agonists and antagonists)
CC of their expression and activity. The antibodies and antagonists are
CC used to down regulate expression and activity of the protein of the
CC invention. The antibodies are also used as diagnostic agents for
CC detecting the presence of the proteins of the invention in samples.
CC The sequences of the invention are also used as vaccines and in gene
CC therapy. The present sequence is human DNA-C-MG.72-1776 cDNA clone
CC encoding novel PRO-C-MG.72 protein of the invention.
XX
SQ Sequence 2366 BP; 788 A; 475 C; 539 G; 563 T; 1 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAD07053 ..

Align seg 1/1 to: AAD07053 from: 1 to: 2366

488 GlyValSerGlnSerLeuSerGly 495
|||||
1321 GGAGTATCTCAAAGCCTTTTCAGGC 1344

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: AAS94303

seq_documentation_block:
ID AAS94303 standard; cDNA; 2488 BP.
XX
AC AAS94303;
XX
XX 13-FEB-2002 (first entry)
DE
XX DNA encoding novel human diagnostic protein #30107.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX

PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG30116.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 30107; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2488 BP; 617 A; 677 C; 671 G; 523 T; 0 other;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAS94303 ..

Align seg 1/1 to: AAS94303 from: 1 to: 2488

370 ValLeuArgArgGluMetArgGln 377
|||||
235 GTCTGCTGCTGCGAATGCTGAG 258

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.AAS27393

seq_documentation_block:
ID AAS27393 standard; cDNA; 2493 BP.
XX
AC AAS27393;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seq ID 428.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;

KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
XX acquired immune deficiency syndrome.
OS Homo sapiens.
XX
PN WO200154733-A1.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01312.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251899.
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
DR P-PSDB; AAU17476.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 428; 880pp; English.
XX
CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AAS27393 ..

Align seg 1/1 to: AAS27393 from: 1 to: 2493

488 GlyValSerGlnSerLeuSerGly 495

|||||
309 GGAGTATCTCAAAGCCTTTCAGGC 332

OM of: US-09-701-711-2 to: GenEmbl.* out_format : pfs
Date: Sep 19, 2002 1:58 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+p2n_model -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09701711/runat_17092002_141401_26890/app_query.fasta_1.882
-DB=GenEmbl -QPMF=fastap -SUFFIX=sepl7std.rge -GAPOP=10.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blowum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs
-NORM_ext=HEARSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09701711 -CGN1_1.8909 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-701-711-2
Query length: 813
Database: GenEmbl.*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 3438.520000

score_list:

Sequence	Strd Orig	zScore	EScore	Len	Documentation
gb_pat:AX009287	+ 4202.00	4796.25	8.5e-259	2442	! AX009287 Sequence 1 from Patent
gb_pat:AX009289	+ 4189.00	4781.39	5.7e-258	2442	! AX009289 Sequence 37 from Patent
gb_pat:AX067462	+ 4189.00	4746.15	5.3e-256	99629	! AX067462 Sequence 37 from Patent
gb_ba:AE004784	+ 1451.50	1636.86	8.2e-83	12469	! AE004784 Pseudomonas aeruginosa
gb_ba:AJ414146	+ 1285.50	1397.41	1.8e-69	210050	! AJ414146 Yersinia pestis str
gb_ba:PL0236920	+ 1258.50	1426.14	4.5e-71	4405	! AJ236920 Photorhabdus luminesc
gb_ba:AL646064	+ 1237.00	1365.25	1.1e-67	201050	! AL646064 Ralstonia solanacea
gb_ba:AL627266	+ 1230.50	1343.66	1.7e-66	268050	! AL627266 Salmonella enterica
gb_ba:AF407013	+ 1212.00	1377.64	2.2e-68	2700	! AF407013 Escherichia coli oute
gb_ba:AE005193	+ 1212.00	1364.60	1.2e-67	10649	! AE005193 Escherichia coli O15
gb_ba:AE000127	+ 1212.00	1361.42	1.8e-67	14877	! AE000127 Escherichia coli K12
gb_ba:AE008705	+ 1212.00	1344.17	3.0e-67	22997	! AE008705 Salmonella typhimuri
gb_ba:ECOTSE	+ 1212.00	1344.17	1.6e-66	91430	! D83536 Escherichia coli genom
gb_ba:ECU70214	+ 1212.00	1341.33	2.4e-66	123171	! U70214 Escherichia coli chro
gb_ba:AP002550	+ 1212.00	1333.48	6.5e-66	281530	! AP002550 Escherichia coli O1
gb_ba:AF120927	+ 1210.00	1376.34	2.6e-68	2433	! AF120927 Shigella flexneri out
gb_ba:AE004297	+ 1203.00	1354.09	4.6e-67	10901	! AE004297 Vibrio cholerae chrc
gb_ba:AE003941	+ 1172.50	1318.55	4.4e-65	11711	! AE003941 Xylella fastidiosa 9
gb_ba:AX155443	+ 1123.50	1277.68	8.3e-63	2379	! AX155443 Sequence 6 from Patent
gb_pat:AX202493	+ 1123.50	1277.68	8.3e-63	2379	! AX202493 Sequence 6 from Patent
gb_pat:AX155439	+ 1115.00	1267.91	2.9e-62	2394	! AX155439 Sequence 2 from Patent
gb_pat:AX202489	+ 1115.00	1267.91	2.9e-62	2394	! AX202489 Sequence 2 from Patent
gb_ba:AE002375	+ 1115.00	1251.44	2.4e-61	13538	! AE002375 Neisseria meningitid
gb_pat:AX043922	+ 1115.00	1220.53	1.3e-59	349980	! AX043922 Sequence 1 from Patent
gb_pat:AX155447	+ 1111.00	1263.34	5.2e-62	2394	! AX155447 Sequence 10 from Patent
gb_pat:AX202497	+ 1111.00	1263.34	5.2e-62	2394	! AX202497 Sequence 10 from Patent
gb_ba:NMA12491	+ 1111.00	1216.21	2.2e-59	340806	! AL162752 Neisseria meningitid
gb_ba:AF021245	+ 1109.00	1261.05	7.0e-62	2394	! AF021245 Neisseria meningitid
gb_ba:NGU81959	+ 1104.50	1255.97	1.3e-61	2379	! U81959 Neisseria gonorrhoeae c
gb_ba:AF329831	+ 1058.50	1203.38	1.1e-58	2382	! AF329831 Haemophilus ducreyi
gb_ba:AE006235	+ 1054.00	1184.42	1.3e-57	10187	! AE006235 Pasteurella multocid
gb_ba:U32773	+ 1053.50	1183.88	1.4e-57	10151	! U32773 Haemophilus influenzae
gb_ba:HIU13961	+ 1052.50	1195.37	3.2e-58	2688	! U13961 Haemophilus influenzae
gb_ba:HIU06833	+ 1052.50	1195.06	3.3e-58	2776	! U60833 Haemophilus influenzae
gb_ba:HIU06834	+ 1052.50	1194.90	3.4e-58	2823	! U60834 Haemophilus influenzae
gb_pat:AX38690	+ 1052.50	1194.49	3.6e-58	2949	! AX38690 Sequence 1 from Patent
gb_pat:AR102148	+ 1052.50	1194.49	3.6e-58	2949	! AR102148 Sequence 1 from Patent
gb_pat:AX38694	+ 1052.50	1194.49	3.6e-58	2950	! AX38694 Sequence 5 from Patent
gb_pat:AR102150	+ 1052.50	1194.49	3.6e-58	2950	! AR102150 Sequence 5 from Patent
gb_pat:AX38692	+ 1052.50	1194.38	3.6e-58	2984	! AX38692 Sequence 3 from Patent

gb_pat:AR102149 + 1052.50 1194.38 3.6e-58 2984 ! AR102149 Sequence 3 from Patent

gb_pat:AX38744 + 1052.50 1194.37 3.6e-58 2987 ! AX38744 Sequence 55 from Patent

gb_pat:AR102157 + 1052.50 1194.37 3.6e-58 2987 ! AR102157 Sequence 55 from Patent

gb_ba:HIU60832 + 1052.50 1189.03 7.2e-58 5236 ! U60832 Haemophilus influenzae

gb_ba:HIU60831 + 1050.00 1185.82 1.1e-57 5438 ! U60831 Haemophilus influenzae

seq_name: gb_pat:AX009287

seq_documentation_block:

LOCUS AX009287 2442 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 1 from Patent WO9963093.

ACCESSION AX009287

VERSION AX009287.1 GI:9996618

KEYWORDS

SOURCE Moraxella catarrhalis.

ORGANISM Moraxella catarrhalis

Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;

REFERENCE 1 (bases 1 to 2442)

AUTHORS Basb027 proteins and genes from moraxella catarrhalis, antigens,

TITLE antibodies, and uses

JOURNAL Patent: WO 9963093-A 1 09-DEC-1999;

FEATURES

LOCATION/Qualifiers

1. .2442

/organism="Moraxella catarrhalis"

/db_xref="taxon:480"

BASE COUNT 737 a 461 c 533 g 711 t

ORIGIN

alignment_scores:

Quality: 4202.00 Length: 813

Ratio: 5.169 Gaps: 0

Percent Similarity: 100.00 Percent Identity: 100.00

alignment_block:

US-09-701-711-2 x AX009287

Align seg 1/1 to: AX009287 from: 1 to: 2442

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17

|||||

1 ATCGGTAAATCATATTTAAAGGTTTTCAGGTCAAGTCAATGACATGGC 50

|||||

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAla 34

|||||

51 TGTCTATGATGTAATGTAATCACTCAATGACAAAGCGCGGATTTTATGGCA 100

|||||

34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50

|||||

101 ATGACATTACCATCACAGGACTACAGCGAGTGACCAATGAAAGCTTACAA 150

|||||

51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67

|||||

151 AGCGTGCTGCGGTTTCGTTGGTGGTCAAGTGGTGAGCAAAACCAAGTTGGC 200

|||||

67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84

|||||

201 TGTATGGTGTCAAGACACTTTATGCAACAGGCAATTTTTCAGATGTGCAAG 250

|||||

84 altYrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100

|||||

251 TCTATCATCAAGAGGCGGTATCATCTATCATAGTAAACGAAAGCGCGTTA 300

|||||

101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117

|||||

301 ATCGCTCAGATTATTTTGGAGGCAATCGCTTAATTCARAAGAAGTGCT 350

|||||

117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134

|||||

351 ACAAGAAGGGCTAAAAAATGCTGGCTTAGCTGTGGGTCAACCACTAAAC 400

134 lAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150
|||||
401 AGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATCA 450
151 GlnGlyTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG 167
|||||
451 CAAGGCTATTATAATACCGAAATTAAGTGTCAACACAGACGATGCTTGATGG 500
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArg 184
|||||
501 TAATCGTGTAACTTGAATGATGACCTTGTCTGAAGTTAAACCTGCACGGG 550
184 alValAspIleAsnIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
|||||
551 TGGTTCATATTAATATCATTTGGCAATCAGCATTTTAGCGATGCAGATTGG 600
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217
|||||
601 ATTGATGTGCTTGCATTAAGGATAATAAAATCAATCCACTGTCTAAAGC 650
217 aAspArgTyrThrGlnGlyLysLeuValThrSerLeuGluAsnLeuArgA 234
|||||
651 TGACCGTTATACTCAAGAAAGCTGTGACCAAGTTTAGAGATTTCGGTG 700
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250
|||||
701 CTAATATCTCAATGCAAGGTTTGTGCGTTTTCAGATTAAAGATGCTAAG 750
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267
|||||
751 CTTAATATTAAAGAGATAAAACCCGTATCTTTGTTGAGATTTCATTTGCA 800
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
|||||
801 TGAGGTGAGCAATATTCGCTTTGGACAGACACAGTTTTCGGTAAATTTAA 850
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
|||||
851 CTTATACTCAAGCAGAACTTGAGGCCTGCTTAAATTCAAAGCAGAAGAA 900
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrly 317
|||||
901 GGGTTTTCACAGCCATGCTTGAGCAAAACAACAATATCAGTACCAA 950
317 sPheGlyAspAspGlyTyrTyrAlaGlnIleArgProValThrArgI 334
|||||
951 ATTTGTTGACGATGGCTATTATTATGCTCAATCCGCTCTGTAACACGCA 1000
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
|||||
1001 TTAATGATGAAAGTCGTACGGTTGATGTGGAATATTATATGACCCCTGTA 1050
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGl 367
|||||
1051 CACCCCTGTATGTACGCCGTATTAATTTTACAGTAACTTTAAGACCCA 1100
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384
|||||
1101 AGATGAAGTACTCCGTCGTGAGATGGACAACTTGAAGTGGCTGGCAT 1150
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
|||||
1151 CTAATCAAAAAATCCAGCTGTCTGTCGACGCTTGATGCGGACTGGGTTT 1200
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspGl 417
|||||
1201 TTTAAACATGTTTACCGTTGATCTCGTCAGTACCACTACCCCTGATCA 1250
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434
|||||
1251 GGTGATGTAAATTTGTGTTGTTGAAGAACCACTTCAGGATCATCAACCA 1300

434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
|||||
1301 TCGCAGCAGGCTACTCTCAAGTGGTGGTAACTTTTCAATTTGATGTT 1350
451 SerGlnAsnAspPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
|||||
1351 TCTCAAAATAAATTTATGGGTACAGTAAGCAGCTCAATGCTTCGGTTTC 1400
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
|||||
1401 TCGCTCTGAGACCCGTGAGGTGTATAGTTTGGGTATGACCAACCCCAT 1450
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
|||||
1451 TTACCGTAAATGGCGTCTCGCAAGCTTGAGTGGCTACTATCGTAAAC 1500
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGl 517
|||||
1501 AAGTATGATAACAAGAACATTAGTAATATGACTTGTATCTTATGGTGG 1550
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
|||||
1551 CTCATTAAAGCTATGGATATCCAAATGATGAAATCAACGCATAAGCTTG 1600
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
|||||
1601 GTCTGAATGCTGACAAATACCAAGCTTCATGGCGGTCGTTTTATGGCATT 1650
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
|||||
1651 AGTAATGTCAAGCAGCTGATGGCAGATGGTGCAAAATTCAGTGGATAA 1700
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIleL 584
|||||
1701 TAATGCAATTCCTGATTTAAGCATGATTACACACCTACAAATGCCATT 1750
584 euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
|||||
1751 TGGGTGGAAATATTCAAGCTAGATCGCCCTGATTTCACAAACCAAGGC 1800
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGl 617
|||||
1801 ATGAGTCAATTCGTAGATTGACGGTTGGTTTGGTGATAAAACCTCATCA 1850
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysServ 634
|||||
1851 AAAAGTGGTTTATCAAGCAATATCTATCGCCCATTTTCAAAAAATCAG 1900
634 alLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
|||||
1901 TCTTGGTGGATACGCCAAAGTTAGGCTATGGCAATAATTTACCATTTAT 1950
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
1951 GAAATTTCTATGACAGCGGCTATGGTTCGGTTCGGGTATGATCAATC 2000
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||
2001 CTCCTTGGGTCCACGCTCACAAAGCTATTGACAGCTCGTCGTGGTCAAC 2050
684 lnThrThrLeuGlyGluValValGlyAsnAlaLeuAlaThrPheGly 700
|||||
2051 AAACCACTAGGAGAGGTTGTTGGTGGTAAATGCTTTGGCAACTTTTCGCG 2100
701 SerGluLeuIleLeuProLeuProPheLysGlyAspTyrIleAspGlnVa 717
|||||
2101 AGTAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATAGATCAGGT 2150
717 lArgProValIlePheIleGluGlyGlnValPheAspThrThrGlyM 734
|||||
2151 GGTCCAGTATATTCATTAGGGGGGCTCAGGTTTTTTGTATACAACAGGTA 2200
734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750

```
|||||
2201 TGGATAACCAACCAATGATTAAACCAATTTAAAGACCCACAAGCAACA 2250
751 AlaGluClnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLys 767
2251 GCTGAACAAATGCAAGAGCAGCAATCGCCGCTACTAACCCCAAGATAA 2300
767 sGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784
2301 ACAGTTGCGTTATAGTCTGCTGCTTGGTGCAACTTGGTATACGCCCATG 2350
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
2351 GTCCTTTATCTATTAGTATGCGCAAGCCATTGAATAAAAAACAAATGAT 2400
801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
2401 CAGACCGATACGATACAGTTCCAGATTGGTAGTGCTCTTT 2439
```

seq_name: gb_pat:AX009289

seq_documentation_block:

LOCUS AX009289 2442 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 3 from Patent WO9963093.

ACCESSION AX009289

VERSION AX009289.1 GI:9996619

KEYWORDS Moraxella catarrhalis.

SOURCE Moraxella catarrhalis.

ORGANISM Moraxella catarrhalis; gamma subdivision; Moraxellaceae;

Bacteria; Proteobacteria;

Moraxella.

1 (bases 1 to 2442)

Vinals-Bassols,C.

Basb027 proteins and genes from moraxella catarrhalis, antigens,

antibodies, and uses

JOURNAL Patent: WO 9963093-A 3 09-DEC-1999;

SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)

FEATURES Location/Qualifiers

source

1..2442

/organism="Moraxella catarrhalis"

/db_xref="taxon:480"

BASE COUNT 735 a 461 c 535 g 711 t

ORIGIN

alignment_scores:

Quality: 4189.00 Length: 813

Ratio: 5.159 Gaps: 0

Percent Similarity: 99.877 Percent Identity: 99.754

alignment_block:

US-09-701-711-2 x AX009289

..

Align seg 1/1 to: AX009289 from: 1 to: 2442

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17

1 ATCCGTAATTCATATTAAAGGTTTCAGTCAATGCAATGGC 50

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaA 34

51 TGTCATGATGGTAATGCAACTCATGCACAAGCGCGGATTTATGGCAA 100

34 sAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50

101 ATGACATTCGCCATCAGAGGACTACAGCGAGTGACCATTTAAAGCTTACA 150

51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67

151 AGCGTCTCGCTTTCGCTTGGGTCAAGTGGTGAGCGACGACAGTTGGC 200

67 aspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValcInV 84

```
201 TGATGCTGTCAAAGCACTTTATGCACAGGCAATTTTTCAGATGTGCAAG 250
84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100
251 TCATCATCAAGAAGGCGGTATCATCTATCAGGTACCGAAAGGCCGTTA 300
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117
301 ATCGCTGAGATTAAATTTTGGGCAATCGCTTAATTCAAAAGAAAGTCT 350
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134
351 ACAAGAGGGCTAAAAAATGCTGGCTTAGCTGGGTCAACCACTAAAC 400
134 lnaLarThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150
401 AGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATATCA 450
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGl 167
451 CAAGGCTATTATAATACCGAAATTTACTGTCAAACAGACGATGCTTGATG 500
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184
501 TAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAACCTGCACGG 550
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
551 TGGTGTATTAATATATCATTTGGCAATCAGCATTTTACGATGCAGATTG 600
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217
601 ATTGATGCTGCTCGGATTAAGGATAATAAAATCAATCCACTGCTAAAGC 650
217 aspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234
651 TGACCGCTATACTCAAGAAAGGTGTTGACGAGTTTAGAGAAATTTGGGTG 700
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250
701 CTAATATCTCAATGCGGTTTGTGCGTTTGTGAGATTAAAGATGCTAAG 750
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267
751 CTTAATATTAAATGAAGATAAAACCGTATCTTGTTCAGATTTCATTGCA 800
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
801 TGAAGGTGAGCAATATCGCTTTGGACAGACACAGATTTTGGGTAATTTAA 850
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
851 CTTATACTCAAGCAGAACTTTGAGGCACTGCTTAAATTCAAAGCAGAGAA 900
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
901 GGGTTTCAACAGCCATGCTTGAGCAACAACAACAATATCAGTACCAA 950
317 sPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334
951 ATTTGGTGACGATGGCTATTATTATGCTCAAAATCCGTCCTGTACACGCA 1000
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
1001 TTAATGATGAAGTGTGCTGACGTTGATGTGGAAATATATATGACCCCTGTA 1050
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGl 367
1051 CACCTCTCTATGTACGCCGTATTAAATTTACAGGTAACCTTTAAGACCCA 1100
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384
1101 AGATGAAGTACTCCGTCGTGAGATGCACAACCTTGAAGGTGCGTTGGCAT 1150
```

```
384  erAsnGlnIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
1151 CTAATCAAAAAATCAGCTGCTCGTGACCGCTTGATCGGACTGGGTTT 1200
401  PheLysHisValThrValAspThrArgProValProAsnSerProAspGI 417
1201 TTTAAACATGTTACCGTTGATACCTCGTCAGTACCACTACCTGATCA 1250
417  nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434
1251 GGTGATGATAAATTTGGTGTGAAGAACAACTTCAGGATCATCAACCA 1300
434  leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
1301 TCGCAGCAGCCCTACTCTCAAGTGGTGGGTAACTTTCAATTTGATGTT 1350
451  SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
1351 TCTCAAAATAACTTTATGGGTACAGTAAGCACGTCATGCTTCGTTTTTC 1400
467  rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
1401 TCGCTCTGAGACCCGTCAGGTGTATAGTTGGGTATGACCAACCCATACT 1450
484  heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
1451 TTACCGGTAATGGCGTCTGCCAAGCTTGAGTGGCTACTATCGTAAACC 1500
501  LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGI 517
1501 AAGTATGATAACAAGAACATTAAGTATATGATGATGATGATGATGATG 1550
517  ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
1551 CCAATTAAGCTATGATATCCATGATGATGATGATGATGATGATGATG 1600
534  lYleuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
1601 GTCTGAATGCTGACAATACCAAGCTTATGCGCGTCTGTTTATGGCATT 1650
551  SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
1651 AGTAATGTCAAGCAGCTGATGTCAGATGTCGCAAAATCAAGTGGATAA 1700
567  nAsnGlyIleProAspPheLysHisAspTyrThrTyrAsnAlaIleL 584
1701 TAATGGCATCTCTGATTTTAAGCATGATTACACAACTACATGCCATTT 1750
584  euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
1751 TGGGTGGAATATTCAAGTCTAGATCGCCCTGTATTTCCAAACCCCAAGC 1800
601  MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGI 617
1801 ATGAGTCATCTCTGATGATTGACGGTGGTGGTGGTGGTGGTGGTGGT 1850
617  nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerV 634
1851 AAAAGTGGTTTATCAAGCAATATCTATCGCCCATTTTATCAAAAAATCAG 1900
634  alLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
1901 TCTTGGCGTGATACGCAAGTTAGGCTATGGCAATAATTTACCATTTTAT 1950
651  GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
1951 GAAATTTCTATGACGCGGCTATGTTGGTTCGGTTCGGTTCGGTTCGGT 2000
667  rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
2001 CTCTTTGGGTCCACGCTCACAGCCTATTGACAGCTCGTGGTGGTCAAC 2050
```

```
684  InThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
2051 AAACCACACTAGGAGAGTTGTTGCTGGTAAATGCTTTGGCAACTTTTCGC 2100
701  SerGluLeuIleLeuProLeuProPheLysGlyAspTyrPheAspGlnVa 717
2101 AGTAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATAGATCAGGT 2150
717  IargProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734
2151 GCGTCCAGTGATATTCATTCATGAGCGCGTCAGGTTTTGATACAAAGGTA 2200
734  etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
2201 TGGATAAAACAAACCATTTGATTTAACCAATTTAAAGACCCCAAGCAACA 2250
751  AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLy 767
2251 GCTGAACAAAATGCAAAAGCAGCCCAATCGCCGCTACTAAACCAAGATAA 2300
767  sGlnLeuArgTyrSerAlaGlyValGlyAlaThrTyrTyrThrProIleG 784
2301 ACAGTTGCGTTATAGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2350
784  lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
2351 GTCCCTTTATCTATTAGCTATGCCAAGCCATTGAATAAAAAACAATAATG 2400
801  GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
2401 CAGACCGATACGGTACAGTCCAGATTGGTAGTGCTTT 2439
seq_name: gb_pat:AX067462
seq_documentation_block:
LOCUS AX067462 99629 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 37 from Patent WO0078968.
ACCESSION AX067462
VERSION AX067462.1 GI:12545082
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
AUTHORS 1 (bases 1 to 99629)
Lagace,R.E., Patterson,C. and Berg,K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 37 28-DEC-2000;
Incyte Genomics, Inc. (US)
FEATURES
source location/Qualifiers
1..99629
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
78467
/note="a or g or c or t, unknown, or other"
BASE COUNT 29233 a 19222 c 21909 g 29264 t 1 others
ORIGIN
alignment_scores:
Quality: 4189.00 Length: 813
Ratio: 5.159 Gaps: 0
Percent Similarity: 99.877 Percent Identity: 99.754
alignment_block:
US-09-701-711-2 x AX067462
Align seq 1/1 to: AX067462 from: 1 to: 99629
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17
ATCGCTAATTCATATTTAAAGGTTTTTCAGGTCAGTCAATGCAATGGC 66182
```

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaA 34
|||||
66183 TGTCTATGATGTAATGTCACAACTCATGACACAGGGCGGATTTATGGCAA 66232
34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50
|||||
66233 ATGACATTTGCCATCACAGGACTACAGGAGTGACCAATGAAAGCTTACAA 66282
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67
|||||
66283 AGCGTGTGCGTTTCGTTGGTCAAGTGGTGAGCGAACACAGTTGGC 66332
67 aspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84
|||||
66333 TGATGGTGTCAAAAGCACTTATGCAACAGGCAATTTTTCAGATGTGCAAG 66382
84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100
|||||
66383 TCTATCATCAAGAAGGCGGTATCATCTATCAGGTAACCGAAGGCGCGTTA 66432
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117
|||||
66433 ATCGCTGAGATTAATTTTGAGGCAATCGCTTAATTCCAAAGAGGTCT 66482
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134
|||||
66483 ACAAGAAGGGCTAAAAAATGCTGGCTAGCTGGTGCAGTCAACCACTAAAC 66532
134 lnaLanThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150
|||||
66533 AAGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATATCA 66582
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGl 167
|||||
66583 CAAGGCTATTAATAACCGAATTAATCTGTCAAACACAGCATGCTTGATGG 66632
167 YAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184
|||||
66633 TAATCGTGTGAAGCTTGATATGACCTTTGCTGAAGGTAAACCTGCACGGG 66682
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
|||||
66683 TGGTGATATTAATATATCATTTGGCAATCACCATTTTAGCGATGCGAGATTG 66732
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217
|||||
66733 ATTGATGCTTCGGATTAAGGATATAAATCAATCCACTGCTAAAGC 66782
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234
|||||
66783 TGACCGTTTACTCAAGAAAGCTGGTGACCACTTTAGAGAAATTTGCGTG 66832
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250
|||||
66833 CTAATATCTCAATGACAGGTTTGTGCGTTTGGAGATTAAGATGCTAAG 66882
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267
|||||
66883 CTTAATTAATGAAGATATAAACCGTATCTTTGTTGAGATTTCAATGCA 66932
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
|||||
66933 TGAAGGTGAGCAATATCGCTTTGGACAGACACAGATTTTGGGTAATTTAA 66982
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
|||||
66983 CTTATATCTCAAGCAGAACCTTGAGGCACTGCTTAAATTTCAAGCAGAGAA 67032
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
|||||
67033 GGGTTTCAACAGCCATGCTTGAGCAACACACAAATATCATGATACCAA 67082
317 sPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334
|||||

67083 ATTTGGTGACGATGGCTATTATTATGCTCAAAATCCGTCCTGTAACACGCA 67132
334 leAsnAspGluSerArgThrValAspValGluTyrIleAspProVal 350
|||||
67133 TTAATGATGAAAGTCGTACGGTTGATGTGGAATATATATTGACCCCTGTA 67182
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGl 367
|||||
67183 CACCCTGCTATGTACGCCGTATTAATTTTACAGGTAACTTTTAAGACCCA 67232
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384
|||||
67233 AGATGAAGTACTCCGTCGTGAGATGCGACAACCTTGAAGGTGCGTTGGCAT 67282
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
|||||
67283 CTAATCAAAAAATCCAGCTCTCGTGCACGCTTGATGCGGACTGGGTTT 67332
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspGl 417
|||||
67333 TTTAAACATGTTACCGTTGATATCTGTCAGTACCCCAACCTCACCTGATCA 67382
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434
|||||
67383 GGTGATGTAATTTTGTGCTTGAAGAACAACCTTTCAGGATCATCAACCA 67432
434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
|||||
67433 TCCACAGCGCTACTCTCAAAAGTGGTGTGTAACCTTTTCAATTTGATGTT 67482
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
|||||
67483 TCTCAAAATAACTTTATGGGTACAGGTAAACGCTCAATGCTTCGTTTTC 67532
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
|||||
67533 TCGCTCTGACACCCGTCGAGGTATAGTTTGGGTATGACCAACCCATACT 67582
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
|||||
67583 TTACCCGTAATGGCGTCTCGCAAGCTTGAGTGGCTACTATCGTAAACCC 67632
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGl 517
|||||
67633 AAGTATGATTAACAAGAACATTAATTAATTAATTAATTAATTAATTAATTA 67682
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
|||||
67683 CTCATTTAAGCTATGGATATCCAATTTGATGAAAAATCAACGCATAAGCTTTG 67732
534 lylLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
|||||
67733 GTCGTAATGCTGACAATACCAAGCTTCATGGCGGTCGTTTATGGGCATT 67782
551 SerAsnValLysGlnLeuMetAlaAspGlyLysIleGlnValAspAs 567
|||||
67783 AGTAATGCTCAAGCAGCTGATGGCAGATGGTGGCAAAATTCAAAGTGGATAA 67832
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIle 584
|||||
67833 TAATGGCATTCCTGATTTTAAGCATGATTACACAACCTACAATGCCATTT 67882
584 euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
|||||
67883 TGGGGTGGAAATTAATCAAGTCTAGATCGCCCTGTATTTTCCAAACCCAGGC 67932
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGl 617
|||||
67933 ATGAGTCATTCGTAGATTTTGACGGTTGGTTTGGTGATAAAACTCATCA 67982
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerV 634
|||||

```

67983 AAAAGTGGTTATCAAGCAATATCATTCGCCCATTTATCAAAAAATCAG 68032
634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnLeuProPheTyr 650
|||||
68033 TCTTCGGTGGATACCCCAAGTAGGCTATGCAATAATTTACCATTTAT 68082
|||||
651 GluAsnPheTyrAlaGlyClyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
68083 GAAAAATTTCTATGACGCGGCTATGTTGCGTTCGTGCTATGATCAATC 68132
|||||
667 tSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgargGlyGlnG 684
|||||
68133 CTCATTGGGTGCACCTCACAGCCTATTTCACAGCTCGTGTGTCACAC 68182
|||||
684 InThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
|||||
68183 AAACACACTAGGAGAGGTTGTTGGTGTATGCTTTGGCACTTTCGCGC 68232
|||||
701 SerGluLeuLeuLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717
|||||
68233 AGTAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGCATAGATCAGGT 68282
|||||
717 IArgProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734
|||||
68283 CGCTCCAGTGATATTCATTAGGCGCGTCAGGTTTTTGATACACAGGTA 68332
|||||
734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
|||||
68333 TGGATAAACAACCATTTGATTTAACCCATTTAAAGACCCACAACACA 68382
|||||
751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLys 767
|||||
68383 GCTGAACAAAAATGCAAAAGCAGCCCAATCGCCGCTACTTAACCCAAATAA 68432
|||||
767 scLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784
|||||
68433 ACAGTTCGCTTATAGTGTGCTGGTGGTGCACTTGGTATAGCCCATGTG 68482
|||||
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
|||||
68483 GTCCTTTATCTATTAGCTATGCAAGCCATTGAATAAAAAACAATAATCAT 68532
|||||
801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
68533 CAGACCGATACGATACAGTTCAGATTCGATGCTGCTTT 68571
|||||
seq_name: gb_ba:AE004784

seq_documentation_block:
LOCUS AE004784 12469 bp DNA linear BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 345 of the complete
genome.
ACCESSION AE004784 AE004091
VERSION AE004784.1 GI:9949799
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 12469)
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
20437337
JOURNAL
MEDLINE
2 (bases 1 to 12469)
Stover,C.K., Pham,X.-O.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,

```

```

Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Sailer,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
Location/Qualifiers
1..12469
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
complement(106..3627)
/gene="dnaE"
/note="PA3640"
complement(106..3627)
/gene="dnaE"
/codon_start=1
/transl_table=11
/product="DNA polymerase III, alpha chain"
/protein_id="AAG07028.1"
/db_xref="GI:9949800"
/translation="MTYSVHLRLHTEFSIVDGLVRVKPLAKAVAGLGMPAVATDQS
NMCSLVFKTAMGAKIPICGADIWLASREDDPLSRSLSLAMNAKGYRNLTELISR
GWSQGRNGEIIIRSDWKEAAEGLIALSAAKEGEIGHALLDGEAAEALKEHMEV
FPERFYVEQRTSRVNDHEHAAVALASRCNAPLATNDVRFIKQDEAEHFRVCI
GEGRTLDDPRRPTYSDQOYKSPAEWELFSDLPEALENTVEIAKRCNEVOLGYFI
LPDPTPNGMGIDDLRHASFEGLEERLEVLVPKPTPDYEAQRQVVDRLNFEELDII
QMGPFGFLIYMDFIKAKNNGVPGVGRSGAGSLVAVYKLTIDLDPLAYDLLEPFI
LNERISMPDFDVFDCMEGRDVRDVIYDAAYGRNAVSIITFTMAKAVVRDVARVQ
KSYGLADRLSRMPEFVGMTLDKAYEQEEMLRDFLKSDEAAEIWEMALKLEITRG
TKHAGGVIAPTKLTDFSPACDEEGGLVTFQDKDVEAAGLVKFDLGLRLTII
KWAMEIINREQAQKGLPEVNIIDPLDDKPTYSLQKAEATAVFOLESRGMEKELIKL
KPDCEDLIALVALFRPGPLOGSMVDVFINRHGRAELSYPHDPDYAGLEPVLKPTY
GIILYQEWQIAQVWAGTYLGGADNRRAMKKKPEEMAKQRGEGIECKNGNIDAD
LAGNIDFLKFAGYFNKSHSAAYGLVSTQATWLKTHFPAPFMAAVLTADMTQDKV
VTLIECRHMKLRIVAPVNDSEFRVDDGGRIVYGLGAIKGVGPGVEAITECHAE
GGPNTLFDFCDVRDLKIRKRTLEALIRAGALDRGLPHYHDELKAYQATVDLNRVL
LAAMEAIAQAEQTASRSHSDHMDLFGVFAEPADVYANHRKVKELTKLERLKGKD
TLGLYLTGHPIDEEYEGEVRFRQIVELKPARDTQTVAGLIVNLVRMKNKGDKMGF
VILDRSGRIEASLSEAPAAQSLQTDALVYVEGEVSQDDPSSGGLRLRAKRVMSLE
EARTGLAESLRMKLHADLLKGRWLWELFNHRGSCPTITDITTSADAKALLQFGES
WRVDPADDLTIQALRDQFGRDNVFLNYR"
/gene="PA3641"
complement(3757..5172)
/gene="PA3641"
complement(3757..5172)
/codon_start=1
/transl_table=11
/product="probable amino acid permease"
/protein_id="AAG07029.1"
/db_xref="GI:9949801"
/translation="MQEPTQVVDLLNGLWGLKILWLLVGCGLYFTRVLGLIQFRHF
GHTFSVLKGRSDSDSGISFQALCTSLAARGVTGNVAGVAITLGGPGAVFWMAI
ALVGMATGEVATLAQLKIRDDKQGRFGPGPAYMEKGLGARWMLFSLFIATFGF
VFNVOANTITGAMOGAFGVTWISVIIVITAIIFGRLSIRAFSELSAVPFPMAVA
YLLAYGIITLNLSELPGLVMVVKSAFGWHEAAAGGIGAAIILNGFKRGLFSNEAGM
SAPNAASATPPYPHPASOGYVOMAGVFIDTLICTASRAIILLAGQPEGELMLVQN
ALTASQGDWGRFVLAVIILFFAFTSIVANFYAENCLVFLFNHLHGLLIFRLILVLA
VMFGALASLPFWVNLADVSMGLMAITNLIAILLLSNLAIKADYNAQRKAGKLPTFD
ASQFPVQOKLEFGIWDERRA"
complement(5389..5994)
/gene="rnhB"
/note="PA3642"
complement(5389..5994)
/gene="rnhB"
/codon_start=1
/transl_table=11
/product="ribonuclease HII"
/protein_id="AAG07030.1"
/db_xref="GI:9949802"
/translation="MQLGLDFNLVDELVAGVDEVGKPGICGPGVVTAAVILDPSPILG

```

```

LNDKSKLSEARREALFEETREKALAWCIARAEEVEIDRLNITLHATMLAMORAVEGLSV
IPRLIDGNRCPKIAPVPCAPVVGKDSOVPATAAAILAKYSROREMYELDRVYPGYG
MAGHKGYTFAVHLEALSRLGTPPIHRRSFAPVRELLDVSVQ"
/sequence="5994. .7130)
/gene="lpxB"
/complement(5994. .7130)
/gene="lpxB"
/codon_start=1
/transl_table=11
/product="lipid A-disaccharide synthase"
/protein_id="AAG07031.1"
/db_xref="GI:9949803"
/translation="MADGLRALVAGEASGDTLGSGLMOALRARHPDIEFTIGVGPRM
EAGLSYFPERLSVGLVEVLGRLPELLRRKRLLITLIEARDVNMIGIDAPDFTL
GVEHKLQAGLRTVHYSPSWAMQKRLVLRACDLMLALFEARFYEEHGVFVR
FVGHPLANTIPQDARAARALGPLADQVLAIMPGRGGEVGLKGLDFTQORLL
VERGLRPLPCASAREQIEQMLQGREPLPLTLDDGASHEALAACDAVLITASGTAT
LEALILKPMVAVYRAGLTIRILKRLVKSPYISLPLNLAGRLVPELIQDAATPOAL
AATLSPLDDGSOQVEFFDAIHRALRQDASQAQAEAVLQVERR"
/complement(7134. .7910)
/gene="lpxA"
/complement(7134. .7910)
/gene="lpxA"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine acyltransferase"
/protein_id="AAG07032.1"
/db_xref="GI:9949804"
/translation="MSLIDPRAIDPSARLAADVGVGWSIVGAEEIGEVTIGPVH
VLKGPTKIGKNNRIYQFSSVEDTDLAYKGEPTRLVIGDHNVIREGVYIHRGTWQDR
AETTGIDHNLIMAYAHIGHDSVIGNHCILVNNALAGHVHDDWAILSGYTLVHQYCR
IGASFSFGMSGAIGKDVPAIVYVFGNPAEASMFNFGMRGRGFSSEALHARLTKVY
YRQGTVEEALAEAEASAAQFPAVAFVFRDSTQSAVRGTR"
/complement(7907. .8347)
/gene="fab2"
/complement(7907. .8347)
/gene="fab2"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine acyltransferase"
/protein_id="AAG07033.1"
/db_xref="GI:9949805"
/translation="MMDINEIREYLPHPYPELLVDVVVELDIEGKRIRAYKNVSNIEP
FNGHFPHPHMPGVLLIEAQAAGILGFRMLDVKPADGLIYFVGSKLRFROPVL
PGDQLHAKFISVKRSIWKFDCHATVDDBKPVCSAEIICAERKL"
/complement(8393. .9454)
/gene="lpxD"
/complement(8393. .9454)
/gene="lpxD"
/codon_start=1
/transl_table=11
/product="UDP-3-O-[3-hydroxylauroyl]"
/protein_id="AAG07034.1"
/db_xref="GI:9949806"
/translation="MMSTLSYTLGQLAHVAEVEGDDADLPLOGLATLQEAQPLSF
LANPOYRKLPESRAGAVLLAADAGFAGTALVVANPILAYASLSHLDFRRKPAAG
IHPTAIAADAEVPSVAGAYAVIESGARIGAGVIGARSYVIGEGWLAAPR
VTLVDHVTIGARVSIQSGAVIGEGGFANEGVWQKIAQIGGVYIGDDVEIGANTTI
DRGALSDFLIGNVKVLDNIMAHNVQIGDHTMAACVIGISAKIGRCHMLAGGVGL
VGHTEICNVFTGTMVYTRSTIEPGSYSSGTAMPQAAEWKKSARIQLDDMARRLO
QLERLAAVTSGDASSDA"
/complement(9454. .9960)
/gene="PA3647"
/complement(9454. .9960)
/gene="PA3647"
/codon_start=1
/transl_table=11
/product="probable outer membrane protein precursor"
/protein_id="AAG07035.1"
```

```

/db_xref="GI:9949807"
/translation="MRKTFQVILITAAIIPAASFAEMKIAVLNOMALLESDAAKQV
AYDAKKFCPQNLKNLKNLERAKADQDLKLVNKGSKMSQDREKALDFKQARDFOFQV
SKELNESKAAADRMLKKLPKLDQAVEETIRKGGYDMVIERGAVVVRPKQYDITRQV
TERMNQLR"
/complement(10011. .12404)
/gene="PA3648"
/complement(10011. .12404)
/gene="PA3648"
/codon_start=1
/transl_table=11
/product="probable outer membrane protein"
/protein_id="AAG07036.1"
/db_xref="GI:9949808"
/translation="MKRELLPALLSALMTAEVAESFTVSDIRVNGLQVRSAGSVFAA
LPLNVEITDDQALVQATKSLRFFQDIQROGNDVLTVTVERPSISSIEIEGNK
AISKEDLLKGLQSLAGEIPORATLEGRNELQVYVAQGRYSAEIAEVIPOPRN
RVALKINNEGTVAAISHINVVGTQVSEEDITDLFELKTNLWFFKNDKYAREKL
SGDLERLASYLDRGYINMDIASTQVSTPDKKHVITVINEGEKYTRDVKLTGDL
KVPEEVRLLLVQKQVFSRKVMTTSDLTRRLNGEYTPANTYNGVPEAHDHDKTV
SVTFVDFGKRAYVNRINFRNKTDEVLRRMRQMEGWSASTYLIDQSKARLERLG
YFKEVNETPAVPTDDQVDVNVISVEEQSPGSSITASVFAQAGLILGSI SONNFI
TGKNSIGLITREYQTRNFQVDPYTWVDSGLYNAFYRTDYDELDVDVASYSVN
SLGAGMSIGYPISETSRITGLSVQDQIDTGRYTVDEIYDFLDKEDGNTNFKASIG
WSESTLNKGLVATRGHSQSLTLETTLPQSDLSGRYKIDYRGQVFAPLTDNRTFRFTEL
GYGDYGSTERLPPFVYAGGFNSVRGFKDSTLGRSPSPSVARNPDGTPMKNOGPD
KGRYTDPDQDPEAFGCGNLLITGGABELLPLPFVKDQRLRTVLFWDVGSTEDTDCPK
TTTNCDDGKIDNLASSVGLTWITALGPLSLATPIKPPDNAETQVQFSLGQTF"
BASE COUNT 2144 a 3895 c 4089 g 2341 t
ORIGIN
alignment_scores:
Quality: 1451.50 Length: 835
Ratio: 2.460 Gaps: 13
Percent Similarity: 70.659 Percent Identity: 36.886
alignment_block:
US-09-701-711-2 x AE004784/rev ..
Align seg 1/1 to reverse of: AE004784 from: 1 to: 12469
7 LysGlyPheGlnValSerAlaMetThrMetAlaValMetMetValMetSe 23
12401 AAAGCGTTTCGTACCGCGCTG...CTGTCAGCGCTGATGATCGCGA 12355
23 rThrHisAlaGlnAlaAlaaspPheMetAlaAsnAspIleThrIleThrG 40
12354 GGTCACGCGAGTCC...TTCACTGTTCCGATATCGGGTCAATG 12311
40 lYLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheArg 56
12310 GCCTGCAGCGGTGTCGCGCGTAGCGTGTTCGCGCGCTGCGCTGAAC 12361
57 LeuGlyGlnValValSerGluAsnGlnLeuAlaAlaaspGlyValLysAlaLe 73
12260 GTCCGCGAGACCATCGACACCGCGCTGTCAGGCGCCCTCCCTCCCT 12211
73 uTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyA 90
12210 GTTCAACACCGCGTTCCTCAGGACATCCAGCTCGCGCGGTGATGGCAACG 12161
90 rgIleIleTyrGlnValThrGluArgProLeuIleAlaGluIleAsnPhe 106
12160 TGGTGGTAGTCACTGTAGTCGAGCGCGCTCGATCTCCAGCATCGAGATC 12111
107 GluGlyAsnArgLeuIleProLysGluGlyGlnGlnGluGlyLysAs 123
12110 GAGGCGCAACAGCGCATCTCCAGGAAGACCTGCTCAAGGCGCTGAAACA 12061
123 nAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetI 140
12060 GTCCGCGCTGCGCGAGCGGAGATCTTCCAGCGCGCCACCCCTCGAAGCG 12011
```

```
140 leGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThr 156
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12010 TCGCAACAGAGCTGCAACCCAGCTAGCTGGCCCGGCTACTCGGCC 11961
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
157 GluIleThrValLysGlnThrMetLeuaspGlyAsnArgValLysLeuAs 173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11960 GAGATCAACGGCGAGTATCCCGCAGCGCCGCAACCGCTCGCGTGAA 11911
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
173 pMetThrPheAlaGluGlyLysProAlaArgValValAspIleAsnIleI 190
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11910 GATCAACATCAACAGGACCGCTGCCCGGATTTCCCATCAACAGTGG 11861
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
190 leGlyAsnGlnHisPheSerAspAlaAspLeuIleaspValLeuAlaIle 206
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11860 TGGGCAACACCGCTTCTCCGAAGAAGACCTGACCGACCTGTTCGAAC 11811
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
207 Lys...AspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThrG1 222
||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
11810 AAGACCACCAACTGGCTGCTGTTCTTCAAGAAGACGACACAAGTACGCCG 11761
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
222 nGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnA 239
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11760 CGAAAGCTCTCCGGGACCTCGAGCGCTCGCTCTCTATTACCTGGACC 11711
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 laGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnGlu 255
||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
11710 GCGGCTACATCAACATGGATATCGCTCCACCCAGGTATCCATCACCCCG 11661
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
256 AspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGlnTy 272
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11660 GACAAGAAGCAGCTGACATCAACCGCTCAACATCAACGAGGTGAGAAGTA 11611
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
272 rArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaAg 289
| | | | | : | | | | | : | | | | | : | | | | | : | | | | |
11610 CACCATCCGGGAGCTGAAGCTACCGGCGACCTGAAGGTCCCGGAAGAGG 11561
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 luLeuGluAlaLeuLeuLysPheLysAlaGluGlyPheSerGlnAla 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11560 AAGTGAAGCGCTGCTGCTGTGTACAGAAGGCCAGGTGTTCTCGCGCAAG 11511
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
306 MetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspG1 322
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11510 GTGATGACCACCACTCCGACCTTAATACCCGCTCGCTGGGCAACGAGG 11461
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
322 yTyrTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerA 339
||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
11460 CTACACCTTCGCCAAGCTCAACGGCTGCCGGAAGCCCATGACGAGACA 11411
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 rgThrValAspValGluTyrTyrIleAspProValHisProValTyrVal 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11410 AGACCGTCTCGGTGACCTTGTGCTGGACCGCGCAAGCGCGCTACGTC 11361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
356 ArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuAr 372
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11360 AACCGCATCAACTTCGCGGCAACACCAAGACCGCAAGACGAAGTGTGG 11311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
372 gArgGluMetArgGlnLeuGlyAlaLeuAlaSerAsnGlnLysIleG 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11310 CCGGAAATCGCCAGATGAAGCGGCTGGGCTCCACCTACCTGATCG 11261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 lnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisValThr 405
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11260 ACCAGTCCAAGGCGCTCTGGAGCGCTCGGCTACTTCAAGGAAGTCAAC 11211
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
406 ValAspThrArgProValProAsnSerProAspGlnValAspValAsnPh 422
||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
11210 GTCGAGACCCCGCGGTTCGCGCAGCCGACGACGAGTCCAGCTCAACTA 11161
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
422 eValValGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrS 439
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11160 CAGCGTGAAGAGACAACCGTCCGCTCGATCACCGCAGCGCTCGGCTCG 11111
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
439 erGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPh 455
||| ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:
```



```

NTKDFGTTLDGVTVKQAGTANIAELKLGVDQVNSQLNLWNGNIGQQVGNKGYSETSVV
LGKYNF"
/ gene      complement(12049. .13929)
/ CDS       /gene="yp01005"
/           complement(12049. .13929)
/           /gene="yp01005"
/ note="Similar to neighbouring CDSs yp01007 and yp01006.
Similar to Salmonella typhimurium leucine-rich repeat
protein SLP TR:O9XCV2 (EMBL:AF127079) (765 aa) fasta
scores: E(): 0. 39.0% id in 644 aa, and to Shigella
flexneri 65.4 kDa antigen IpaH 4.5 SW:IPA4_SHIP4 (P18009)
(574 aa) fasta scores: E(): 0. 40.4% id in 589 aa"
/ codon_start=1
/ transl_table=11
/ product="putative antigenic leucine-rich repeat protein"
/ protein_id="CAC89848.1"
/ db_xref="GI:15979074"
/ translation="MNSLSTNSVMPNIEPDREIHSARTSTAALTPADYVIAIWEKWE
SEAPGDEQRRAVECMKDLKNTYHLCLRDNLASLPDILPPCNELDMGNKLTVE
LPATLPDNLQKLNASFQNLRTLPDTPASLLSNVYGNELERPELPEGLKELDVND
NESLQLPNRPPELPSLGIASCGLTLPNLSKRLDADSNOLRTLPDTPSLNL
SVNSNOLTPQPELPSLGIASCGLTLPNLSKRLDADSNOLRTLPDTPSLNL
PQCEILLGNPSTSLQVQLHLRINPYQGRINWSELDNPPASLRNIVATWLP
EQNRLAGDANITEANSAFVFLHRLATQNNANPEFKQOIAWLQLQADSPFL
REGTFLAQASATCEDRITLTHNDQKAVMLHEVEKGYDEKLPMLARGEMFRLE
QLENIAREKVIILKTLNVNSVDIEVYLAYQVKLLSLQSSVKNMRFFGVSHVTAD
DLLSAETRVKTAENQDSRWLSQSPKSVVQRIEPEPYAAAEKQVHALENTYDPKL
AAELAAANGMGDGDANRIVKRNDELGMELDMALTHEVLSKAGSSLLDNLWMEYLI
SP"
/ complement(13099. .13164)
/ gene="yp01005"
/ note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 9.30, E-value 44"
/ gene="yp01005"
/ note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 12.80, E-value 8.4"
/ complement(13294. .13353)
/ gene="yp01005"
/ note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 10.40, E-value 30"
/ complement(13468. .13539)
/ gene="yp01005"
/ note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 17.30, E-value 0.36"
/ complement(13543. .13602)
/ gene="yp01005"
/ note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 12.90, E-value 7.8"
/ complement(14492. .15367)
/ gene="yp01006"
/ note="Similar to neighbouring CDSs yp01005 and yp01007.
Similar to regions of Salmonella typhimurium secreted
protein h2 TR:Q9RPH0 (EMBL:AF160727) (788 aa) fasta
scores: E(): 2.5e-24, 41.9% id in 272 aa, and to regions
of Yersinia pestis outer membrane protein YopM TR:O68701
(EMBL:AF053946) (409 aa) fasta scores: E(): 7e-23, 40.5%
id in 259 aa"
/ codon_start=1
/ transl_table=11
/ product="putative antigenic leucine-rich repeat protein"
/ protein_id="CAC89849.1"
/ db_xref="GI:15979075"
/ translation="MYLSNLTNSVMPNIEPDREIHADREPAATATLPADYHAIWEKWE
NDPRTVAGEORGQAVAKMKELENNAERLNSLDLSTLPDTPPCNELMILCNLTVE
LPTLPDNLQTLKASYNQNLRTLPDTPASLLSKVHMNELERPELPEGLKELDVGC
NTSLQPSRPPVLESIDJNSLQSLPDLAHLRQNCNCLDGNPLSPSTLLAL
NLVAYNOITLPELNPGLSLRCIYTYEYNQLSQLPDLAHLRQNCNCLDGNPLSPSTLLAL
LRLSTKPNYQGPRI"
/ complement(14657. .14728)
/ gene="yp01006"

```

```

/ note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 15.40, E-value 1.3"
/ complement(14732. .14791)
/ gene="yp01006"
/ note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 9.30, E-value 43"
/ complement(14981. .15040)
/ gene="yp01006"
/ note="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 11.30, E-value 22"

alignment_scores:
  Quality: 1265.50      Length: 826
  Ratio: 2.256         Gaps: 14
  Percent Similarity: 67.918  Percent Identity: 35.472

alignment_block:
US-09-701-711-2 x AJ414146

Align seg 1/1 to: AJ414146 from: 1 to: 210050

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23
74360 CGCATAACAACGATGGCGATGAAAAAGTTGCTCATAGCGTCGCTGCTGT 74409
24 .....ThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleT 37
74410 TGGCAGCGCCACCGTATACGGTGCAGCGGGTTCGTAGTGAACGATATTC 74459
37 hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53
74460 ATTTCGAAGGACTGCAACGGGTGCGCGTGGTCTGCTTAAATATG 74509
54 ProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVa 70
74510 CCGGTTCGGTAGGCGATACCGTCAGTGATGATGATCGGTAAACACTAT 74559
70 llysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG 87
74560 CCGTTCGGTGTGTTGTCACAGCAACTTCGAGGACGTCGCGCTCGCGG 74609
87 InGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGlu 103
74610 ATGGTAATACGCTGATTGTTCAAGTCAAGAACGCCGCCACGATGCCAGC 74659
104 IleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGl 120
74660 ATCACTTTCGGGTAAATAACGGCGTGAAGAAGATATGCTCAACAGAA 74709
120 yLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrV 137
74710 TCTGGAAGCCTCTGCGTCCGGTGGCGGCTTGGACCGAACCACTA 74759
137 alGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 153
74760 TATCAACATCTGAAAAGACGCTCGAAGATTTCTATTACAGTGTGGGTAA 74809
154 TyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVa 170
74810 TACACGGCTTCAGTGAAGCGTGTGTTAGCCCATTACCACGTAATCGTGT 74859
170 llysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspI 187
74860 TGACCTGAATTTGGCTTTTACCGAAGGTGTTTCTGCAAAAATTCAGCAG 74909
187 leAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 203
74910 TTAATATCGTAGCTATCACAGTTTCACCACCTGATGAGCTTATCTCACGC 74959
204 LeuAlaIleLysAsp.....AsnLysIleAsnProLeuSe 215
74960 TTCCAACCTCGGTGACGAGGTGCTTGGTGGAAACGTAGTC..... 74998

```



```

76729 ATCCAGTTTAAACATTGGTAAACTTGG 76756
seq_name: gb_ba:PLU236920

seq documentation block:
LOCUS   PLU236920               4405 bp    DNA         linear    BCT 02-AUG-1999
DEFINITION   Photorhabdus luminescens yael (partial), fira (partial), oma and ompH genes.
ACCESSION   AJ236920
VERSION     GI:5689864
KEYWORDS    fira gene; oma gene; ompH gene; outer membrane antigen; periplasmic protein; yael gene.
SOURCE      Photorhabdus luminescens.
ORGANISM    Photorhabdus luminescens
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Photorhabdus.
REFERENCE   1 (bases 1 to 4405)
            Chatonnet-Marton,P.I., Givaudan,A., Lanois,A. and Boemare,N.E.
            Photorhabdus luminescens genomic region homologous to 4.0 minute Escherichia coli region promotes pleiotropic phenotypes
            Unpublished
            Submitted (12-FEB-1999) Chatonnet-Marton P.I., Laboratoire de Pathologie Comparee, Institut National de la Recherche Agronomique, Universite Montpellier II, C.C.101, Pl. E. Bataillon, Montpellier, 34095, FRANCE
FEATURES             Location/Qualifiers
     source          1..4405
                     /organism="Photorhabdus luminescens"
                     /strain="Hm"
                     /db_xref="taxon:29488"
                     /dev_stage="phase I variant"
     gene            1..581
                     /gene="yael"
     CDS             1..581
                     /codon_start=1
                     /transl_table=1
                     /protein_id="CAB51928.1"
                     /db_xref="GI:5689865"
                     /translation="MEKVPGSAAEKAGLQKDRIVKVGSGQIEDVHWHTTSTFVSNNPN
                     VPLEUSDRAHHIISLMSPEVROQSGGRKVFAGVELRIVPLADSYKIKVQIQGPESA
                     MYQAGDKTVMRLTIVSMIGKLVGVGDIKNNLSGPISIAKAGVSADSGVLVYLFPLA
                     LISVNLGINLIPVLVDGHLFLFIEKIKGGPVSERVQDFSYRIGAMILVLLMGLA
                     LFNDFSRP"
                     596..703
                     /gene="oma"
     RBS             join(696..703,716..3109)
                     /gene="oma"
                     716..3109
                     /gene="putative"
                     /note="putative"
                     /codon_start=1
                     /transl_table=1
                     /product="outer membrane antigen"
                     /protein_id="CAB51929.1"
                     /db_xref="GI:5689866"
                     /translation="MAMKGLLIASLLFGSAAAYCAGDFVVQDIIHFEGLRQVAVGAALL
                     NMPVRVGTSDSDIGRTIHAFPLGFNFDVVRDGNLTIVQKRPRTIASITFSGN
                     KSVKDKMLKQLEASHVRVCGALDRTMLSNIERGLEDFYYSVGKYNASKVAVYPLPR
                     NRVDLKVFAAGVSAKIQQINIVGNKSFSSDELLNRPOLRDDVPWNRLADQYKQKQ
                     RTDGLAURSYLDGRYARFNIDSTQVSLTPDKKGIYVFINMTEGQYKISGIDLNGN
                     MAGYQSEITKLAALPEPSLYNGTQVTMENDIKNLGRYGAYPRVMTQPEINDDQKT
                     VKLHVINDAGNRVYVRKIRFSGNDTTKDSLRMRQMERAWLGSDLVPELGRNRL
                     GYFTVDVETQRIIPGSDQDVVYKVKERNVTGSLNFGVGGTSGVSFOIGAQQDNWL
                     GTGNAGVYNASKNDSYFSTFDPTVTINGVSLGRVFYNDFRADDAELSGYTNQS
                     YGISGFLGPTVNNNSLNFLNLYIHNSLDMLPQVAMWRYLRSKGEKPOLSKAEKA
                     DDFALTMGPTVNNIDRFPFPGSVKSTLNGKVTIPGSDNEFYKVTLDTSAYTINDDR
                     TWVILGRSLRYGDLGGKGLPFFYENFYAGSSTVRGFSNNIGKPAIYLDKDGSPKK
                     ESPSDAYKGNMVAASLELTPPTPLDSYSNSVTSFTFIDSGTVMZDWDNDVYMK
                     SKGIPDYKGNINVSAGIALQWMSPLGVLFVSYAKPKIDYEGDRSEQEFOQNLGKTV"

```

68 pGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValt 85
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
898 TACCATTATGCACCTATTTGCTACTGGTAACTTTCAACACGCTTCTGTC 947
85 yrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIle 101
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
948 TGCCTGATGGCAATACACATTTATTTCTAGGTAAAGAGCGCGCACTATT 997
102 AlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuG1 118
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
998 GCTAGCATCATCTTCTCCGGCAATAAATCCGGTGAAGATGACATGTTGAA 1047
118 nGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnA 135
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1048 ACAGAACCTTGAGGCATCTCTGTTGCGGTGCGAGAACCCCTTGACCGCA 1097
135 laThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGln 151
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1098 CGATGCTGCTCAATATCAAGAGAGGTTGGAAGATTTCTATTACAGTGTG 1147
152 GlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAs 168
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1148 GGCAATACACGGCGAGGTAAAGCGGTGTTACGCGCTTCCACGCTAA 1197
168 nArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValV 185
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1198 CCGTGTCCATTTAAACCTGGTTTTGCTGAAGGTGTTTCGGCAAAATCC 1247
185 alAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIle 201
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1248 AGCAAAATTAACATGTCGGTAAATAAATCATTTCTCTGATGAGTTATTG 1297
202 AspValLeuAlaIleLysAspAsnLysIleAsnPro.....LeuSe 215
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1298 ANTGGTTCCAACTGAGAGATGATGTG.....CCGTGGTGGAAATCTGAC 1341
215 rLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnL 232
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1342 CGCTGATCAGAAATATCAGAAACAACTGCTGCTGACCTTGAAGCAT 1391
232 euArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysasp 248
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1392 TGGCAGTTTTTACCTTGATCGCGGTATGCGCGTTTCAACATGATTTCG 1441
249 AlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSe 265
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1442 ACTCAGGTAGCTTGACGCCAGATAAAAGGTATTTATGTCACGATAAA 1491
265 rLeuHisGluGlyGluIntYrArgPheGlyGlnThrGlnPheLeuGlyA 282
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1492 TATGACTGAGGGGATCAATAAATAATATCCGGTATTGACCTGAACGGTA 1541
282 snLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAla 298
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1542 ATATGGCGGGTTATCAGTCAGAAATTTACTAAGCTGGCTGCCATTCAGCCT 1591
299 GluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSe 315
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1592 GGTCTCTGTATACGGGACTCAGGTGACTAAATGGAAATGATATCAA 1641
315 rThrLysPheGlyAspAspGlyTyrTyrAlaGlnIleArgProValT 332
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1642 AAACCTGCTGCTGCTATGCTTATGCTTATCCAGGATGATGACGCAAC 1691
332 hrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleAsp 348
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1692 CTGAATTAATGATCAGATAAGACATAAAGTAAACATGTCATGTAATTTGAT 1741
349 ProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPheLy 365
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1742 GCAGGCAACCGCTTCTATGTCGGTAAATCGTTTTTCTGTTAATGACAC 1791

365 sThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaL 382
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1792 CACTAAAGATTCGCTGTCGCTGGAATCGTTCAGATGGAAGAGCAT 1841
382 euAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThr 398
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1842 GGTTAGGAGGTGATTTAGTTGAACCTAGTAAAGAGCGCTTTAACCGTTTG 1891
399 GlyPhePheLysHisValThrValAspThrArgProValProAsnSerPr 415
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1892 GGTATTTCGAAACGGTAGATGTTGAAACTCAGCGTATACCGAGCGCC 1941
415 oAspGlnValAspValAsnPheValValIleGluGlnProSerGlySers 432
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1942 TGATCAGCTCGATGTCGTTTATAAAGTTAAAGAACGTAATACCGGTCTCTC 1991
432 erThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPhe 448
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
1992 TGAACCTTTGGTGTGTTTGTACAGAAAGCGCGTTAGTTTCCAAATT 2041
449 AspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSe 465
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2042 GCGCTCAACAGGATACTGGTTGGGTACAGGTAACGCTGTGGAAATTAA 2091
465 rPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnP 482
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2092 TGCCAGTAAGAACGATTTTCAACTTACGACAGCTCTCCTTCACCTGATC 2141
482 roTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArg 498
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2142 CATACTTTACGATTAATGGTGTGAGCCCTGGTGTGCGAGTATTCTACAAC 2191
499 LysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTy 515
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2192 GATTTACAGCGCGATGCTGCTGAATTTACCGGTTACACTAACCAAGCTA 2241
515 rGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIles 532
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2242 TGGTATAAGCGGTTTCCCTGCTTCCCGATTAATGAAAAATAACTCGCTGA 2291
532 erPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMet 548
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2292 ATTTTGGCTTAAATATATTATCACTCTCTG...TCTGATATGCTCTCT 2338
549 GlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnVa 565
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2339 CAGGTAGCAATGTGCGTTATCTGAGGTCTATGGCGAGAACCTGATCT 2388
565 lAspAsnAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnA 582
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2389 TGAAGTAAAGCC.....GAATTCAAAGCAGAT.....GATTTTGCTC 2426
582 laIleLeuGlyTyrPAsnTyrSerSerLeuAspArgProValPheProThr 598
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2427 TCACATATGGATGGACATATAACAACCTCGACCGTGGTTCTTCCCAACA 2476
599 GlnGlyMetSerHisSerValasp.....LeuThrValGlyPheGlyAs 613
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2477 TCAGGTGGAAGTCGACTCTGAATGGTAAAGTGACTATTCTGGTTCGGA 2526
613 pLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPhe. 629
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2527 TAACGAATTTTACAAAGTGACTCTTGTACTCTTCGCTATTATCAATTA 2576
630IleLysLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGly 644
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2577 ATGATGATCGTACTTGGTGTGATTTGGGACGCTGCGGTTTAGGGTATGGG 2626
645 Asn.....AsnLeuProPheTyrGluAsnPheTyrAlaG1 656
::: |||||:::|||||:::|||||:::|||||:::|||||:::|||||
2627 GATGGATTAGCGGTAAGAGTTGCCATTCATGAAAACTTCTATGCAGG 2676
656 yGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgS 673


```

/misc_feature
/transl_table=11
/product="ribosome recycling factor"
/protein_id="CAD08677.1"
/db_xref="GI:16501500"
/translacion="MISDIRKDAEVRMEKCEAFKQISKVRTGRASPSLLDGIVVEY
YGTPTPLQRLASVTVDISRLKINFDKSMGPAVERKAIMASDLGLNPSAGTDIRVPL
PPLTEERKDLTKIVRGEAEOARVAVNRDRANDKVKALLKDKAKAISEDDEDRRSQEEV
OKMTDAIKKVDALADKEALMQF"
3086..3580
/gene="STV0242"
/notes="Pfam match to entry PF01765 RRF, Ribosome recycling
factor, score 339.90, E-value 2.9e-98"
3731..4927
/gene="dxr"
/notes="STV0243; yaem"
3731..4927
/gene="dxr"
/EC_number="1.1.1.-"
/notes="Similar to Escherichia coli 1-deoxy-D-xylose
5-phosphate reductoisomerase dxr SW.DXR_ECOLI (P45568;
P77209) (398 aa) fasta scores: E(): 0, 88.9% id in 397 aa
orthologue of E. coli yaem (DXR_ECOLI); Fasta hit to
DXR_ECOLI (398 aa), 89% identity in 397 aa overlap"
/codon_start=1
/transl_table=11
/product="hypotheical protein in frr 3'region"
/protein_id="CAD08678.1"
/db_xref="GI:16501501"
/translacion="MKQLTILGSTGICSTLDVVHNPDSFRVIALVAGKNVMAE
QCLEFSRYAVMDTSSAQKIMLQOHGSRTEVLSSQQAACEMALDEGVHMAAIV
GAAGLPTLAIRAGKTILLANKESLVTGCRLEFMDVKRNSNARLLPVDSEHNAIFOSL
POSTOHLNGYADLEONGVTSILLTGSGGPERETPCMDLAAMTPDOACRHPNWSMGRI
SVSATMNGKLEYIEARWLFNASARQMEVLIHPQSVIHSWRYQDSVLAQLEGPD
RTPAHTMAMPNRYSCAQEDFCCKLSALTFSDPYORYFCLKLAMEAFEGQQAQTA
LNANEITVFAAQOIRFTDIDAGLNLAVLERMDLHEPASVDDVLQVDAAREVARKQ
VIRLSR"
4064..4087
/gene="dxr"
/notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
5240..5998
/gene="STV0244"
5240..5998
/gene="STV0244"
/notes="Orthologue of E. coli UPPS_ECOLI; Fasta hit to
UPPS_ECOLI (253 aa), 94% identity in 252 aa overlap"
/codon_start=1
/transl_table=11
/product="undecaprenyl pyrophosphate synthetase"
/protein_id="CAD08679.1"
/db_xref="GI:16501502"
/translacion="MLSATQPVSENLPAHGRHVAIIMDGNRWAKQKIRAFGHKA
GAKSVRRVAFNANGIDALTLYAFSSNNWRNPAQEVSAIMELFVWALDSEVSLHRH
NVLRIIGDLSRNSRLQERLRKSEALTHTNTGLTLNIAANYGRWDIVOGVRLAEL
VOAGVLRPDQIDERLQQQICMHELAPVDLIVRTGGEHRISNELLWQIAELYFTDV
LWPDFDQDEFGALHAFANRRPFGGTEPGDDKA"
5306..5971
/gene="STV0244"
/notes="Pfam match to entry PF01255 UPP_synthetase,
Putative undecaprenyl diphosphate synthase, score 492.70,
E-value 2.8e-144"
5804..5857
/gene="STV0244"
/notes="PS01066 Undecaprenyl pyrophosphate synthetase
family signature"
6011..6868
/gene="STV0245"
/notes="cdsa"
6011..6868
/gene="STV0245"
/notes="Fasta hit to YNBB_ECOLI (298 aa), 33% identity in
298 aa overlap
Orthologue of E. coli cdsA (CDSA_ECOLI); Fasta hit to

```

```

alignment_scores:
  Quality: 1220.50      Length: 845
  Ratio: 2.211         Gaps: 15
  Percent Similarity: 65.325 Percent Identity: 33.728

alignment_block:
US-09-701-711-2 x AL627266 ..

Align seg 1/1 to: AL627266 from: 1 to: 268050

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8252 CGCATAATAACGATGCGGATGAAAAAGTTGCTCATAGCGTCGCTGTT 8301
24 .....ThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleT 37
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8302 CAGCAGCGCCACCGTATACGGTGTGAAGGGTTCGTGTGAAGGACATTC 8351
37 hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8352 ATTTCGAGGCGCTGCGGAGTCCGGTGTGGTGGCTCTCCTCAGTATG 8401
54 PropheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVa 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8402 CCGGTGCGCACAGCGACACGGTTAATGATGAAGATATCAGTAACACCAT 8451
70 llyAsnLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG 87
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8452 TCGCGCTCTGTTCCGACCGGCACTTTGAGGACGTCGCGCTCCGCGCG 8501
87 InGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGlu 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8502 ATGGTAAATACCTTCTGGTTCAGTAAAAAGACGTCGACCATTCGACG 8551
104 IleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGl 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8552 ATCACTTTCTCCGGGAACAAGTCGGTGAAGATGACATGCTCAAGCAAA 8601
120 vLeuLysAsnAlaGlyLeuAlaValGlyGlnProLysGlnAlaThrV 137
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8602 CCTCGAAGCGTCTGGGTACGTTGGCGAGTCTCTGGATCGCACACGC 8651
137 alGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8652 TGTCTGATATCGAAAAAGTCTCGAAGACTTTTACTATAGCGTAGTAAA 8701
154 TyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVa 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8702 TACAGCGCAGGTTAAGCGGTGCTCAGCGCGCTACCGCGCAACCGTGT 8751
170 llysLeuAspMetThrPheAlaGluGlyLysProAlaArgValAlaSpi 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8752 TCGATCTAAACTGGTGTTCAGGAGGCGGTATCGCGGAAGATCAACAGA 8801
187 leAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8802 TCAACATCGTCGGTAACCATGCTTTAGCACCGCAAGAACTCATTTCTCAT 8851
204 LeuAlaIleLysAsp.....AsnLysIleAsnProLeuSe 215
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8852 TTCAGGTCGCTGACGAGTGGCGGTGGTGAACGTCGTC..... 8890
215 rLysAlaAspArg...TyrThrGlnGluLysLeuValThrSerLeuGluA 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8891 ....GCGGATCGTAAATACAGAAACAGAAAGCTGCGCGGCGACCTTGA 8936
231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLys 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8937 CCTTGGCGAGCTACTATCTGGATCGCGGCTACCGCGCTTTCAATATTGAC 8986
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluI 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
8987 TCCACGCAAGTGAGTCTGACGCGGGATAAGAAAGGATCTACATCACCGT 9036

```

```

264 eSerLeuHisGluGluGlnThrArgPheGlyGlnThrGlnPheLeuG 281
    : : : : : | | | | | : : : : : : : : : : : : : : : |
9037 CAATATCACCAAGGCGATCAGTACAAAGCTTCCGGGGTTCCAGTGAGCG 9086

281 lysLeuThrThrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLys 297
    : : : : : | | | | | : : : : : | | | | | : : : : :
9087 GTAACTGGCCGCGACATTCGCGCTGAAATTGAGAACTGACGAAATTTAA 9136

298 AlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnI 314
    | | | | | : : : : : : : : : : : : : : : : : : : :
9137 CCGGGCGAGCTCTATACGGCCCAAGATGACCAAAATGGAAGATGACAT 9186

314 eSerThrLysPheGlyAspAspGlyThrThrAlaGlnAlaGlnProv 331
    : : : : : | | | | | : : : : : | | | | | : : : : :
9187 CAAAAGCTTCTGGTGGCTGATGCTACGCTTACCGCGGTGCAGTCAC 9236

331 alThrArgIleAsnAspGluSerArgThrValAspValGluThrThrIle 347
    : : : : : | | | | | : : : : : | | | | | : : : : :
9237 AGCTTGAATTAACGATGCTGACAAACCGTAAACTGCGCGTGAACGTC 9286

348 AspProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPh 364
    | | | | | : : : : : | | | | | : : : : : | | | | |
9287 GATCGGGCAACCGTTCTACGTGCGTAAGATCCGCTTGAAGGCAACGA 9336

364 eLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyA 381
    : : : : : | | | | | : : : : : | | | | | : : : : :
9337 TACCTCAAAAGATTCCGTTCTGCGCGTGAAATCGCCAGATGGAAGGGG 9386

381 laLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArg 397
    | | | | | : : : : : : : : : : : : : : : : : : : :
9387 CGTGGCTGGCGACGACCTGGTTGACCAAGGGAAGAGCGTCTGAACCCG 9436

398 ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414
    | | | | | : : : : : | | | | | : : : : : | | | | |
9437 CTGCGGTTCTTTGAACCGTCGACCGCACCGACCGCGGCTTCCGGGAG 9486

414 rProAspGlnValAspValAsnPheValValGluGlnGlnProSerGlyS 431
    | | | | | : : : : : : : : : : : : : : : : : : : :
9487 CCGGATCAGTTGATGTGGTGACAAAGGTGAAAGCGGTAACACCGGTA 9536

431 eSerThrIleAlaAlaGlyThrSerGlnSerGlyGlyValThrPheGln 447
    | | | | | : : : : : | | | | | : : : : : | | | | |
9537 GCTCAACTTCGGTATCGGTACGGTACAGAAAGCGCGGTGAGCTTCCAG 9586

448 PheAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAl 464
    | | | | | : : : : : | | | | | : : : : : | | | | |
9587 GCCGGGTTTCAGCAGCACTAGTACGCGGTATTCGCGTATTCGCGGGAT 9636

464 aSerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrA 481
    : : : : : | | | | | : : : : : | | | | | : : : : :
9637 TAACGGTACCAAAATGACTACCAAGCTTACAGCGAACTGTCCGTTACTA 9686

481 snProThrPheThrValAsnGlnValSerGlnSerLeuSerGlyThrThr 497
    | | | | | : : : : : | | | | | : : : : : | | | | |
9687 ACCATACCTTTACCGTTGCGGGGTGAGCGTGGTGGTGGTATTCCTTCTAT 9736

498 ArgLysThrLysThrAspAsnLysAsnIleSerAsnTyrValLeuAspSe 514
    : : : : : | | | | | : : : : : | | | | | : : : : :
9737 AACGACTTTGAACGGATGATCGCGACCTGTCGCACTATACCAACAGAG 9786

514 rTyrGlyGlySerLeuSerThrGlyThrProIleAspGluAsnGlnArgI 531
    | | | | | : : : : : | | | | | : : : : : | | | | |
9787 TTACGGTACCGATGTGCGCTTGGCTTCCCATCAACGAGTACACACAGT 9836

531 leSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPhe 547
    : : : : : | | | | | : : : : : | | | | | : : : : :
9837 TCGTGCTGGTTGGCGTATGTCCATTAACAAACTG..... 9871

548 MetGlyIleSerAsnValLys...GlnLeuMetAlaAspGlyGlyLysIl 563
    | | | | | : : : : : | | | | | : : : : : | | | | |
9872 .....TCCAACTGACGCGCGAGATCGCGATGGATGATCTTGA 9912

```

```

563 eGlnValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrThr 580
    : : : : : | | | | | : : : : : | | | | | : : : : :
9913 ATCGATGGGCGACCCATCATCCAGCGATTTGCCGCTGACGACTTCACCT 9962

580 yrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspArgProValPhe 596
    : : : : : | | | | | : : : : : | | | | | : : : : :
9963 TTAAAC.....TACGGCTGACCTATAACAAGCTTGACCGTGGTTATTTC 10006

597 ProThrGlnGly.....MetSerHisSerValAspLeuThrValGlyPh 611
    | | | | | : : : : : : : : : : : : : : : : : : : :
10007 CCACGGACGGCTCGCGAGTCAACCTGACGGTAAAGTGACGATCCAGG 10056

611 eGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgP 628
    : : : : : | | | | | : : : : : | | | | | : : : : :
10057 TTCGACAAACGAATACTACAAAGTGTGCTGGATACCGGACCTATCTGC 10106

628 roPheIleLys.....LysSerValLeuArgGlyTyrAlaLysLeuGly 642
    | | | | | : : : : : | | | | | : : : : : | | | | |
10107 CTATTGATAACGACCAACAGTGGTGGTCTTGGGGCGTACTCGTTGGGT 10156

643 TyrGlyAsnAsnLeu.....ProPheTyrGluAsnPheTy 654
    | | | | | : : : : : | | | | | : : : : : | | | | |
10157 TACGGCGATGGTTTAGCGGCAAGGAATCCCGTCTCTAGAACTCTA 10206

654 rAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyP 671
    | | | | | : : : : : | | | | | : : : : : | | | | |
10207 TCGCGGTGGTTCCAGTACCGTGGTGGTTCAGTCCGAATATATCGGCC 10256

671 roArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnGlnThrThr... 686
    | | | | | : : : : : | | | | | : : : : : | | | | |
10257 CGAAAGCG.....GTCTATAAAATGGCGCCACACTAGTTGG 10294

687 .....Le 687

10295 GACGATGATGATGATTACGAGACTGTACTCAGGAATCAGCGTGTAAATC 10344

687 uGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuI 704
    : : : : : | | | | | : : : : : | | | | | : : : : :
10345 AGATGATCGGTAGGTGGTAAATGCCATGGCTGTCGCCAGCGCTGGAATTA 10394

704 leLeuProLeuProPheLysGlyAsp...TrpIleAspGlnValArgPro 719
    | | | | | : : : : : | | | | | : : : : : | | | | |
10395 TTACCCCGACGCGCTTCATCAGCGAAAGTATGCCAACTCGTCCGTACC 10444

720 ValIlePheIleGluGlyGlnValPheAspThrThr...GlyMetas 735
    : : : : : | | | | | : : : : : | | | | | : : : : :
10445 TCCTTCTTCTGGGATATGGTACCGTGGGATACGAACCTGGGATCCGTC 10494

735 pLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaG 752
    | | | | | : : : : : | | | | | : : : : : | | | | |
10495 ATCTGCGCGCTCAGATGTACCGGATTACAGCGATCCA..... 10531

752 luGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLysGln 768
    : : : : : | | | | | : : : : : | | | | | : : : : :
10532 .....GGCAAC 10537

769 LeuArgTyrSerAlaGlyValGlyAlaThrTriPtyrThrProIleGlyPr 785
    : : : : : | | | | | : : : : : | | | | | : : : : :
10538 ATCCGTATGTCGCGGGTATCGCATTTACATGGATGTCCCATTTGGGCC 10587

785 oLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAspGlnT 802
    | | | | | : : : : : | | | | | : : : : : | | | | |
10588 GTTGGTCTTCTCTACGCCAGCGGTTTAAAAAGTACGATGGAGACAAG 10637

802 hrAspThrValGlnPheGlnIleGlySerValPhe 813
    : : : : : | | | | | : : : : : | | | | | : : : : :
10638 CCAGCAGTTCAGTTTAAATTTGGTAAACCTGG 10672

```

seq_name: gb_ba:AF407013

seq_documentation_block:

LOCUS AF407013

DEFINITION Escherichia coli outer membrane protein (ecfk) gene, complete cds.

2700 bp

DNA

linear

BCT 10-SEP-2001


```
1131 CAAAAGCTTCGCGTCGCTATGGTTATGCTATCCGCGGTACAGTCGA 1180
331 aThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIle 347
1181 TGCCCGAAATTAACGATGCCGCAAAACCGTTAAATTACGTGTGAACGTT 1230
348 AspProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPh 364
1231 GATGCGGGTAACCGTTTCTACGTGCGTAGATCCGTTTGAAGGTAAACA 1280
364 eLysThrGlnAspGluValLeuArgGluMetArgGlnLeuGluGlyA 381
1281 TACCTCGAAAGATGCCGTCCTCGCTCGCGAAATGCGTCAAGTGAAGGTG 1330
381 laLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArg 397
1331 CATGCTGCGGGAGCGATCTGGTCGATCAAGGTAAGGAGCGCTGAATCGT 1380
398 ThrGlyPheLysHisValThrValAspThrArgPProValProAsnSe 414
1381 CTGGGCTCTTTGAACACTGCTCATACCGATACCCAAACGTGTTCCGGGTAG 1430
414 rProAspGlnValAspValAsnPheValValGluGluGlnProSerGlys 431
1431 CCGGACCAAGTTGATGTCGTCTACAAGGTAAGGAGCGCAACACCGGTA 1480
431 eSerThrIleAlaIleGlyTyrSerGlnSerGlyGlyValThrPheGln 447
1481 GCTTCAACTTTGGTATTGTTACGTTACTGAAAGTGGCGTGAGCTTCCAG 1530
448 PheAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAl 464
1531 GCTGTGTGCACGAGTAACCTGGTTAGGTACAGTTATGCTGTTGGTAT 1580
464 sSerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrA 481
1581 CAACGGGACCAAAACGATTACCAACCTATGCTGAACCTGTCGGTAACCA 1630
481 snProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyr 497
1631 ACCCTACTTACCGTAGATGCGGTAAGCGTCGGTGGTGGTCTCTCTAT 1680
498 ArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSe 514
1681 AATGACTTCCAGCGAGATGACCCCGCACCTGTCGACTATACCAACAGAG 1730
514 rTyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgI 531
1731 TTATGGTACAGACGTACGCTTGGGCTTCCGATTAAACGAATATAACTCGC 1780
531 leSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyArgPhe 547
1781 TCGGTGACGCTCGGTGTATGTACATAAATCCCTG..... 1815
548 MetGlyIleSerAsnValLysGlnLeuMetAla..... 558
1816 .....TCCACATGACAGCGCTCAGGTGGATGTCGCGTTATCTGTA 1856
559 AspGlyGlyLysIleGlnValAspAsnGlyIleProAspPheLysH 575
1857 CTCATGGGTGAACATCCGAGCACCTCTGTAGTACAGGATAACAGCTTCAAAA 1906
575 isAspTyrThrThrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeu 591
1907 CGGACGACTTACGTTCAAC.....TATGGTGGACCTATATAACAGCTT 1950
592 AspArgProValPheProThrGlnGly.....MetSerHisSerValas 606
1951 GACCGTGTACTTCCGACAGATGGTTCAAGTGTCAACCTGACCGGTAA 2000
606 pLeuThrValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnG 623
```

```
2001 AGTGACCATTTCTGGATCGGATAACGAATACTACAAAGTCACGTTAGACA 2050
623 lyAsnIleTyrArgProPhe.....IleLysLysSerValLeuArgGly 637
2051 CGCGGACTTATGTGCGCGATCGATGACGATCACAATGGGTGTCTCTGGGG 2100
638 TyrAlaLysLeuGlyTyrGlyAsnAsnLeu.....Proph 649
2101 CGTACCCGCTGGGTATTGTGTATGTTAGCGCGCAAGAGATGCCGTT 2150
649 eTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspG 666
2151 CTACGAGAACTTCTATCGCGGTGTCCAGCACCGTGGCTTCCAGT 2200
666 lnSerSerLeuGlyProArgSer.....GlnAla..... 675
2201 CCAATACCATTTGTCGAAAGCAGTTTACTTCCCGCATCAGGCGCAATAAT 2250
676 .....TyrLeuThrAlaArgArgGlyGlnGlnThrTh 686
2251 TATGATCCGCGACTATGATTACGAATGTGCGACTCAGGACGCGCGAAGA 2300
686 rLeu.....GlyGluValValGlyGlyAsnAlaLeuAlaThrPheG 700
2301 CCTGTGTAATCGGATGATGCTGTAGCGGTAAACGCATGGCGTTGCCA 2350
700 lySerGluLeuIleLeuProLeuProPheLysGlyAsp...TrpIleAsp 715
2351 GCGTCGAGTTTCATCACCCGCGCGCTTATTAGCGATAAGTATGCTAATC 2400
716 GlnValArgProValIlePheIleGlyGlyGlnValPheAspThrTh 732
2401 TCGGTTCTGACTCTCTCTCTGGGATATGGTACCGTTTGGGATACAA 2450
732 rGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnA 749
2451 CTGGGATTCACGCCAATATCTGGATATCCGACTATAGTATGCCA.... 2496
749 laThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGln 765
2496 ..... 2496
766 AspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThr 782
2497 ...AGCAATATCCGATGTCGCGGTATCGCATTAATGGATGCCCC 2543
782 oileGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA 799
2544 ATTGGGCGGTGGTGTCTCTACGCCCGCGTTCAAAAAGTACGATG 2593
799 snAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
2594 GAGCAAGCGCAACAGTTCCAGTTTAACATCGTTAAAAACCTGG 2637
```

seq_name: gb_ba:AE005193

seq_documentation_block:

LOCUS AE005193

DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 17

of 155.

ACCESSION AE005193

VERSION AE005193.1

KEYWORDS GI:12512893

SOURCE Escherichia coli O157:H7 EDL933.

ORGANISM Escherichia coli O157:H7 EDL933

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

REFERENCE 1 (bases 1 to 10649)

AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,

Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,

Posfal, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,

Grotbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamou, S.K.,

Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D., Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K., Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C., Welch, R.A. and Blattner, F.R.
 Direct Submission
 Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 FEATURES
 source
 Location/Qualifiers
 1. 10649
 /organism="Escherichia coli O157:H7 EDL933"
 /strain="EDL933"
 /serotype="O157:H7"
 /db_xref="taxon:155864"
 /note="enterohemorrhagic"
 215. 940
 /gene="rpsB"
 /note="rpsB"
 215. 940
 /gene="rpsB"
 /note="rpsB"
 /function="structural component; Macromolecule synthesis, modification; Ribosomal proteins - synthesis, modification; Ribosome"
 /note="Residues 1 to 241 of 241 are 100.00 pct identical to residues 1 to 241 of 241 from Escherichia coli K-12 strain MGI655: B0169"
 /codon_start=1
 /transl_table=11
 /product="30S ribosomal subunit protein S2"
 /protein_id="AAG54471.1"
 /db_xref="GI:12512894"
 /translation="MATVSMRDMKAGVHGHOTVWPNKMKPFIFGARKNKHIIINE KTVPMNEALAEINKIASRKGIKILFVCTKRAAEAVKDAALSCDQFVNNHRLWGLMT NWTYRQSIKRLKLETSQSDGTFDKLTKEALMRPELEKLSLGGIKDMGLDPA LFVIDADHEHIAKEANNIGIPVFAIVDTNSDDPGVDVFIPIGNDDAIRAVTLILGAVA ATVREGRSODLASQAEESEVEAE"
 1198. 2049
 /gene="tsf"
 /note="tsf"
 1198. 2049
 /gene="tsf"
 /function="factor; Macromolecule synthesis, modification; Proteins - translation and modification"
 /note="Residues 1 to 283 of 283 are 100.00 pct identical to residues 1 to 283 of 283 from Escherichia coli K-12 strain MGI655: B0170"
 /codon_start=1
 /transl_table=11
 /product="protein chain elongation factor EF-Ts"
 /protein_id="AAG54472.1"
 /db_xref="GI:12512895"
 /translation="MAITASLVKELRRTGAGMDCCKKALTEANGDIELAIENMRKS GAIKAAKAGNAADGVKIKIDNGYGIILEVNCQDTEFAKADAGFAADKVLDAVA GKITDVEVLKAEEREAVALVIGENIRINRAALBGDVLGYSQHGARGVLVAAG ADELKVKHAMHAAKSPFIKPEDVSAEYVEKYQVOLDIAMQSGKPKRIAEKMEVG RMRKFTGTEVSLTQPFVMEPSKTVGQLLKEHNAEVTGFIREFVGEIEKVTETDFAAEV AAMSKQS"
 2196. 2921
 /gene="pyrH"
 /note="pyrH"
 2196. 2921
 /gene="pyrH"
 /function="enzyme; Central intermediary metabolism: Nucleotide interconversions"
 /note="Residues 1 to 241 of 241 are 100.00 pct identical to residues 1 to 241 of 241 from Escherichia coli K-12

Strain MGI655: B0171"
 /codon_start=1
 /transl_table=11
 /product="uridylylate kinase"
 /protein_id="AAG54473.1"
 /db_xref="GI:12512896"
 /translation="MATNAKPVYKRIKLLKSGALQGTGFGDASILDMAQEIKEL VELGIQGVVIGGNLFRGAGLAKAGMNRVYGDHGMGLATVMNGLAMRDALHAYVNA RLMSAIPLNGVCDSDYSWAEBASILLRNNRVILSAGTGNPFTTDSACLRLGITEADV LKATKVDGVTADPAKDPATMYEQLTYSVELEKELKVMDLAAFTLARDHKLPIRVF NNNKPCALRRVYMGKEGELITE"
 3071. 3628
 /gene="fir"
 /note="fir"
 3071. 3628
 /gene="fir"
 /function="factor; Macromolecule synthesis, modification; Proteins - translation and modification"
 /note="Residues 1 to 185 of 185 are 100.00 pct identical to residues 1 to 185 of 185 from Escherichia coli K-12 strain MGI655: B0172"
 /codon_start=1
 /transl_table=11
 /product="ribosome releasing factor"
 /protein_id="AAG54474.1"
 /db_xref="GI:12512897"
 /translation="MISDRKDAEVRMKCVAEAFKQISKIRTPGRASPSLLDGIVVEY YGPTPLQLASVTVEDSRCLKINVFDRSMSPAVEAKAIMASDLGLNPSAGSDIRVPL PPTERRKDLTKIVRGEAQARVAVNRVRDANDKVKALLKKEISEDDDRRSQDIDV OKLTDAAIKIEALADKAEELMQF"
 3720. 4916
 /gene="yaeM"
 /note="yaeM"
 3720. 4916
 /gene="yaeM"
 /function="putative transport; Not classified"
 /note="Residues 1 to 398 of 398 are 99.74 pct identical to residues 1 to 398 of 398 from Escherichia coli K-12 Strain MGI655: B0173"
 /codon_start=1
 /transl_table=11
 /product="putative ATP-binding component of a transport system"
 /protein_id="AAG54475.1"
 /db_xref="GI:12512898"
 /translation="MKQLTILGSTIGCTLDVVRHNPEHFRVVALVAGKNVTRMVE QCLEFSPRYAVMDDEASAKLLKTMLOQGSRTVELSQQAACDMALEDVDVDMAAIV GAAGLLPTLAAIRAGKTIILANKESLVTGCLFMDAVKQSKAOLLPVDSHNAIFQSL PQIOLNGLYADLEQNGVSVILLTGSGGPFREPLRDLATWTPDQACRHPNWSMGRKI SVDSATMMNKLEYIEARWLFNASQSEVLIHPQSVIHSNRYQDSVLAQLGEPDM RTPIAHTMWPNRVNSGVKPLDFCKLSALTFAAPDYPCLKLAMEAFEGQAATTA LNAANEITVAFLAQQIRFTDIAALNLSLEKMDPREPCQVDVDSLVSASAREYARKE VMRLAS"
 5102. 5863
 /gene="yaeS"
 /note="yaeS"
 5102. 5863
 /gene="yaeS"
 /function="orf; Unknown function"
 /note="Residues 1 to 253 of 253 are 100.00 pct identical to residues 1 to 253 of 253 from Escherichia coli K-12 strain MGI655: B0174"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAG54476.1"
 /db_xref="GI:12512899"
 /translation="MMLSATQPLSEKLPANHCRRHVAIIMDNGEWAQKQKIRAFGHK AGKSVRRVAFSANNNGIEALTLYAFSENWNRPAQEVSALELFWALDSEVSLRHE HNVRLRIIGDTSRFNSRLQRIKSEALTATGTLTGLTNIAANYGGWDIVOGVQLAE KVQOQNLPDQIDDEMLNQHVCHELAPVDLIVRTGGEHRISNLLMQIAAYAEIFYTD VLPWFDEQDFEGALNAFANRRERFEGTPEGDETA"
 5984. 6733

[illegible]

10170	TATGATCGGACTATCATACGAATGTCGGACTCAGGACGGCGGAAGA	10211
686	rIeu.....GlyGluValValGlyGlyAsnAlaLeuAlaThrPheG	700
10220	CTGTGTAAATCGGATGCTGTAGGCGGTAAACCCATGCGGTGGCA	10269
700	lySerGluLeuIleLeuProLeuProPheLysGlyAsp...TrpIleASP	715
10270	GCCTCGAGTTCATACCCGAGCGCGTTTATTAGCGATAAGTATGCTAAC	10319
716	GlnValArgProValIlePheIleGluGlyGlnValPheAspThrTh	732
10320	TCGGTTCGTACTCTCTCTCGGATATGGTACCGTGGGATACAA	10369
732	rglyMetAspLysGlnThrIleAspLeuThrGlnPhelYsAspProGlnA	749
10370	CTGGGATTCACGCCAATATCTCGGATATCCGAGCTATAGTGATCCA...	10415
749	laThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGln	765
10415	seq_name: gb_ba:AE000127	10415
766	AspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrPr	782
10416	...ACCAATATCGTATGCTCTCGGATATCGCATTAATGATGATCCCC	10462
782	oiledglyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA	799
10463	ATTGGGGCGGTGGTGTCTCTCTACGCCAGCGCTTCAAAAAGTACGATG	10512
799	snAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe	813
10513	GAGACAGCGACAAACAGTCCAGTTTAACATCGGTAAACCTGG	10556
seq_name: gb_ba:AE000127		
seq_documentation_block:		
LOCUS	AE000127	14877 bp DNA linear
DEFINITION	Escherichia coli K12 MG1655 section 17 of 400 of the genome.	
ACCESSION	AE000127	U00096
VERSION	AE000127.1	GI:1786370
KEYWORDS		
SOURCE	Escherichia coli K12.	
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia	
REFERENCE	1 (bases 1 to 14877)	
AUTHORS	Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Mau,B. and Shao,Y.	
TITLE	The complete genome sequence of Escherichia coli K-12	
JOURNAL	Science 277 (5331), 1453-1474 (1997)	
MEDLINE	97426617	
PUBMED	9278503	
REFERENCE	2 (bases 1 to 14877)	
AUTHORS	Blattner,F.R.	
TITLE	Submitted (16-JAN-1997) Guy Plunkett III, Laboratory	
JOURNAL	University of Wisconsin, 445 Henry Mall, Madison, WI	
	Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	
REFERENCE	3 (bases 1 to 14877)	
AUTHORS	Blattner,F.R.	
TITLE	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory	
JOURNAL	University of Wisconsin, 445 Henry Mall, Madison, WI	
	Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	
REFERENCE	4 (bases 1 to 14877)	
AUTHORS	Plunkett,G. III.	
TITLE	Direct Submission	

JOURNAL

Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

COMMENT

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA. 30332 [e-mail: markamber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

Location/Qualifiers

```

1..14877
  /organism="Escherichia coli K12"
  /strain="K12"
  /sub_strain="MG1655"
  /db_xref="taxon:83333"
123..884
  /gene="yaeS"
  /note="b0174"
123..884
  /gene="yaeS"
  /function="orf; Unknown"
  /note="0253; 61 pct identical to 229 residues of an approx.240 aa hypothetical protein Y920_HAEIN SW: P44938"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAC73285.1"
  /db_xref="GI:1786371"
  /translation="MMLSATQPLSEKLPAGCRHVAIIIMDNGRWAKQKIRAFGHK
  AGAKSVRAVFAANNIGTEALTLYAFSSNNRPAQEVSALEMLFVWALDSEVKS LHR
  HNVRLITGDTSRFSNRQERIRKSEALTAGNTGLTLNIAANYGGRWDIVQGVQLAE
  KVQGNLPDQDIDEMLNQHVCMHELAPVDLIVRTGGHRI SNFLWQIAELVFTD
  VLWPDFDQDREGALNAPNRERRGGTEPGEFA"
870..900
  /note="factor Sigma70; predicted +1 start at 195687"
899..927
  /note="factor Sigma70; predicted +1 start at 195714"
956..987
  /note="factor Sigma70; predicted +1 start at 195774"
1005..1754
  /gene="cdsA"
  /EC_number="2.7.7.41"
  /function="enzyme; Fatty acid and phosphatidic acid biosynthesis"
  /note="0249; 100 pct identical to CDSA_ECOLI SW: P06466"
  /codon_start=1
  /transl_table=11
  /product="CDP-diglyceride synthetase"
  /protein_id="AAC73286.1"
  /db_xref="GI:1786372"

```

protein_bind

```

/translation="MLAAEWGQSGFTTRSORVLAIVLCGLLLALMLFLPEYHRNI
HQPLVEISLWASLGHWWIYALLLVLYEYPGSAAIWRNSKILRIFGVLTIVYFPFGMLAL
RAWHYDENHYSGAILWLLYVLLVWGADSGAYMFGKLGKRGKHLKAPKVSCKTQWQGTGG
LATAAIVSWGYGMWANNLDVAPVTLTLCISIVAALASVLGDLTESMPKREAGIKDSGHLI
FGHGILDRIDSLTAAPVPVFACLLLLVFRTL"
1421..1432
  /gene="cdsA"
  /note="No predicted promoter"
  /bound_moiety="PutA predicted site"
1766..3118
  /gene="yaeL"
  /note="b0176"
1766..3118
  /gene="yaeL"
  /function="orf; Unknown"
  /note="0450; 100 pct identical to fragment YAEI_ECOLI
  SW:P37784 but contains 302 additional C-terminal residues"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAC73287.1"
  /db_xref="GI:1786373"
  /translation="MLSFLMDIASFVALGVLTIVHERGHFWARCGVRVRFSGIF
  GKALWRTDKLGTETEVIALPLGGYKMLDERAEVPELHAFNNKSVGQRAAIIA
  AGPVANFIFAYVWLVTIIGVPGVVPVGEIAANSAIAEQIAPGTALKAVDGIETP
  DWDVRLQLVDKIGDESTITVAPFGSDQRDVKDLDRHWAPEPKEDPVSSIGIRPR
  GPQIPELVNQPNASAASKAGIQAQDRIVKVDGQPLTQWTFVMLVRPNQPSLALEI
  ERQGSPLTILIPESKPGNKAIGFVGIPEKVIPLPDEKYKVRQGPFNALVEADTKT
  WOLMKLTVMIGKLITGDVKNLNSGPISIAKGAGMTAELGVVYLYPLFLALISVNLGI
  INLPFLPVLDDGGHLLFLAIEKIKGGPVSERVQDFCYRIGSILLVLLMGLALNDFSRLL
  "
3148..5580
  /gene="yaeR"
  /note="b0177"
3148..5580
  /gene="yaeR"
  /function="orf; Unknown"
  /note="0810; 45 pct identical (29 gaps) to 808 residues of
  an approx. 800 aa protein D151_HAEIN SW: P46024"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAC73288.1"
  /db_xref="GI:1786374"
  /translation="MAMKLLIASLLFSSATVYGAEGFVKDIHPEGIQRVAVGAALL
  SMPVPTGPTVDEDSNTIRALFATGNEEDVRLPDGDTLLVQVKERTIASITFSGN
  KSVDDMLKQLEASGVVGVESLDRITTIADIEKLEDFYISVGKYSASVAVVFLPR
  NRVDLKLVFQGVSAEQIINIVGNHAFITDELISHFQLRDEVPWNVVGDKKYQKQK
  LAGDLETLSYLDRGYARFNIDSTQVSLTPDKGIIYTVVNIITEGDTYKLSGVEYSGN
  LAGSAETEQLTKEPGLYNGTKVTKMEDDIDKLLGRYAYPRVQSMPEINDADKT
  VKLRVNDAGNRVYVRKTRFEGNDTSKDAVLRREMRQMEGAWLGSDLVDQGERLNRL
  GFFETVDTQRPQSPQDVVYKVKERTGSGNFSGIGYCTESVSGVQAGVQDQNW
  GTGAVGLNGKNDQYTAELSVTNPYFTVDGSLGRFLYNDFAADLSDYTNKS
  YGTDVTLGFPINENSRAGLSYHNLSNMQPVAMWRLYLSMEHSTSDQNSFK
  TDDTFNYGWTYNKLDRGYFPTDGRVNLTKGVTIPDGSKYKVTLTATVYPTDQD
  HKWVVLGRTRWYGDGKGPFPYAGSSTVGVQSGNTGPKVYIPHOASNY
  DPDYDECATODGAKDLCKSDDAVGNAMAVASLEFITPTPEISPKYANSVTSFWD
  MGTVDVNDWDSQYSGYDPSNIRMSAGIALQWMSPLGLPVSFAQPKKYDGK
  AEQFQFNIGKTW"
5583..5611
  /note="factor Sigma70; predicted +1 start at 200398"
5702..6187
  /gene="hlpA"
  /note="b0178"
5702..6187
  /gene="hlpA"
  /function="factor; Basic proteins - synthesis, modification"
  /note="0161; 100 pct identical to HLP_A_ECOLI SW:
  P11457; similar to ompH of other bacterial species"
  /codon_start=1
  /transl_table=11
  /product="histone-like protein, located in outer membrane

```

promoter

gene

CDS

or nucleoid"
 /protein_id="AAC73289.1"
 /db_xref="GI:1786375"
 /translation="MKWLLAAGLIGLALATSAQAADKIAIVNMGSLFOOVAOKTGVSN
 TLNEPKFASERLQRMETDLOAKMKKLOSMKAGSDRTKLEKDYMAQRQTFPAQKAQAFQ
 ODRARRSNERGKLVTRIQTAFAVKSANSDIDLVVDANAVAYNSSDKDITADVLKQV
 K"

gene

6191..7216

CDS

/gene="lpxd"

6191..7216

/note="b0179"

/gene="lpxd"

/EC_number="2.3.1.-"

/function="enzyme; Surface polysaccharides and antigens"

/note="o341; 100 pct identical to LPXD_ECOLI SW: P21645

but includes initiator met"

/codon_start=1

/transl_table=11

alignment_scores:

Quality: 1212.00 Length: 848
 Ratio: 2.200 Gaps: 15
 Percent Similarity: 64.976 Percent Identity: 33.844

alignment_block:

US-09-701-711-2 x AE000127 ..

Align seg 1/1 to: AE000127 from: 1 to: 14877

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23
 3136 CGCATTAATACGATGCGGATAAAGTTGCTCATAGCGTCGCTGTT 3185
 24ThrHisAlaGlnAlaAlaaspPheMetAlaAsnAspIleT 37
 3186 TAGCAGCGCCACCGTATACGGTCTGAAGGTTCTGTAGTGAAGATATTC 3235
 37 hrIleThrGluLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53
 3236 ATTTCGAAGGCTTCAGCGTGCCTGCGGTGGTGGCGCCCTCTCAGTAG 3285
 54 ProPheArgLeuGluGlnValValSerGluAsnGlnLeuAlaAspGlyVa 70
 3286 CCGGTGCGCAGCGGACACAGGTATATGATGAAGATATCAGTAATACCAT 3335
 70 lLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG 87
 3336 TCGCGCTCTGTTGTCACGCAACTTTGAGGATGTTGCGCTCCTCGTG 3385
 87 lGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGlu 103
 3386 ATGGTGATACCTCTCTGTTTCAGGTAAGAAAGACGTCGACCATTCGACG 3435
 104 lLeAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
 3436 ATTACTTCTCCGTAACAATACCGTGAAGATGACATGCTGAAGCAAAA 3485
 120 yLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrV 137
 3486 CCTCAGGCTTCTGTTGCGGTGGGCAATCCCTCGATCGCACCA 3535
 137 aGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 153
 3536 TTGCGCATATCAGAAAGCTCGGAAGACTTCTACTACAGCGTCGGTAAA 3585
 154 TyrAsnThrGluThrValLysGlnThrMetLeuAspGlyAsnArgVa 170
 3586 TATAGCGCCAGGTAAGCTGCTGACCCCGCTGCCGCAACCGTGT 3635
 170 lLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValAlaSpI 187
 3636 TGACCTAAAACCTGGTTCAGGAAGGTGTGTCAGCTGAATCCAGCAAA 3685

187 lLeAsnIleGlyAsnGlnHisPheSerAlaAspLeuIleAspVal 203
 3686 TTAACATTGTTGGTAACCATGCTTTTACCACCGCACTGATCTCTCAT 3735
 204 LeuAlaIleLysAsp.....AsnLysIleAsnProLeuSe 215
 3736 TTCCAACCTGCGTGACGAAGTCCGCTGGTGAACGTGGTA..... 3774
 215 rLysAlaAspArg...TyrThrGlnGluLysLeuValThrSerLeuGluA 231
 3775GGCATCTGTAATACCAAGAACAACTGGCGGCGACCTTGAAA 3820
 231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluLys 247
 3821 CCCTCGCAGCTACTATCTGGATCGCGTTATGCCGTTTCAACATCGAC 3870
 248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluI 264
 3871 TCTACCCAGGTCAGTCTGACGCCAGATAAAAAGGTATTTACGTACCGT 3920
 264 eSerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuG 281
 3921 GAACATCACCGAAGCGGATCAGTACAAGCTTCTGCGGTGAAGTCAGCG 3970
 281 lYAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLys 297
 3971 GCAACCTTGGCGGCACCTCCGCTGAATTTAGCAGCTGACTAAGATCGAG 4020
 298 AlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnI 314
 4021 CCGGTGAGCTGTATAACGGCACCAAGTGAAGTGAAGATGACAT 4070
 314 eSerThrLysPheGlyAspAspGlyTyrTyrAlaGlnIleArgProV 331
 4071 CAAAAGCTTCTCGCTCGCTATGTTATGCCATTCGCGCGTACAGTCGA 4120
 331 alThrArgIleAsnAspGluSerArgThrValAspValGluTyrIle 347
 4121 TGCCGGAATTAACGATGCCGAAACCGTTAAATTTACGTGTGAACGTT 4170
 348 AspProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPh 364
 4171 GATGCGGTAAACCTTCTACGTGCTAAGATCCGTTTGAAGGTAACGA 4220
 364 elYsthrGlnAspGluValLeuArgGluMetArgGlnLeuGluGlyA 381
 4221 TACCTCGAAGATGCGCTCGTCCGCAATGCGTCAGATGGAAGGTG 4270
 381 lLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArg 397
 4271 CATGCTGGGAGCGATCGTGCATCAGGGTAAGGCGTCTGAATCGT 4320
 398 ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414
 4321 CTGGGCTCTTTGAAACTGTCGATACCGATACCAACCGTTCGCGGTAG 4370
 414 rProAspGlnValAspValAsnPheValValGluGlnProSerGlyS 431
 4371 CCGGACCAAGTGTGATGCTCTACAAGTAAAGAGCGCAACCGGTA 4420
 431 erSerThrIleAlaAlaGlyTyrSerInsSerGlyGlyValThrPheGln 447
 4421 GTTCAACTTTGGTATGTTAGGTAGTGAAGTGGCGGTGAGCTCCAG 4470
 448 PheAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAl 464
 4471 GCTGTGTGCACGAGGATAACTGGTTAGGTACAGGTTATGCTGTTGGTAT 4520
 464 aserPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrA 481
 4521 CAACGGGACCAAAACGATTACCAACCTATGCTGAACCTGTCGGTAACCA 4570
 481 snProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyr 497

```

|||||
4571 ACCCGTACTTACCGTAGATGGCGTAAAGCTCGTGGTCTCTCTCTAT 4620
498 ArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSe 514
4621 ATGACTTCCAGCAGATGACCGCGACCTGTCGACTATACCAACAAGAG 4670
514 rTyrGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgI 531
4671 TTATGGTACAGACGTGACGTGGCTTCCGATTCCGATTACGAATATACTCGC 4720
531 leSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPhe 547
4721 TCGGTGCGAGTCTGGGTATGTACATAACTCCCTG..... 4755
548 MetGlyIleSerAsnValLysGlnLeuMetAla..... 558
4756 .....TCCAACTAGCAGCTCAGGTGCGATGTGGCGTATCTGTA 4796
559 AspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysH 575
4797 CTCTATGGGTGAACATCCGACGACCTCTCATCAGGATACAGCTTCAAAA 4846
575 iAspTyrThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeu 591
4847 CGACGACTTACGTTCAAC.....TATGGTTGGACCTATACAAAGCTT 4890
592 AspArgProValPheProThrGlnGly.....MetSerHisSerValAs 606
4891 GACCGGTGTTACTTCCGACAGATGGTTCACGTGTCACACTCAACCTCGGTA 4940
606 pLeuThrValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnG 623
4941 AGTGACCATTCCTGGATCGGTAACAACTACAAAGTGACGTTAGACA 4990
623 lYAsnIleTyrArgProPhe.....IleLysLysSerValLeuArgGly 637
4991 CGCGCACTTATGTCGCGATCGATGACGATCACAATAATGGTTGTCGGG 5040
638 TyrAlaLysLeuGlyTyrGlyAsnAsnLeu.....ProPh 649
5041 CTATCCCGTGGGTATGGTATGGTATGGCGCAAGAGATCGCGTT 5090
649 eTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspG 666
5091 CTACGAGAATTCTATGCGCGGTGTTCCAGCACCGTGGCTGCTCCAGT 5140
666 lNserSerLeuGlyProArgSer.....GlnAla..... 675
5141 CCAATACCATTTGTCGGAAGCAGTTACTTCCCGCATCAGGCCAGTAAT 5190
676 .....TyrLeuThrAlaArgArgGlyGlnGlnThrTh 586
5191 TATGATCCGCACTATGATTACGAATGTGCGACTCAGGACGCGCGAAGA 5240
686 rLeu.....GlyGluValValGlyGlyAsnAlaLeuAlaThrPheG 700
5241 CTGTGTAATCGGATGATGCTGAGCGGTACGCCATGCGCGTTGCCA 5290
700 lYserGluLeuLeuLeuProLeuProPheLysGlyAsp...TrpIleAsp 715
5291 GCCTCGAGTTTCATACCCGCGCGCGCTTTATTAGCGATAAGTATGCTAAC 5340
716 GlnValArgProValIlePheIleGluGlyGlnValPheAspThrTh 732
5341 TCGGTTCGACTTCCTCTCTCTGGGATATGGGTACCGTTTGGGATACAAA 5390
732 rGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnA 749
5391 CTGGGATCCAGCCCAATATCTGGATATCCGACATATAGTATGATCA... 5436
749 laThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGln 765

```

```

5436 ..... 5436
766 AspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrPr 782
5437 ...AGCAATATCGTATGCTCGGGTATCGCAATTACATGGATGCCCC 5483
782 olleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA 799
5484 ATTGGGCGGTGGTGTCTCTCTACGCCAGCGCTTCAAAAAGTACGATG 5533
799 snAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
5534 GAGACAAGGCAACAACAGTTCCAGTTTAACTCGGTAAACACCTGG 5577

seq_name: gb_ba:AE008705

seq_documentation_block:
LOCUS AE008705 22997 bp DNA linear BCT 25-OCT-2001
DEFINITION Salmonella typhimurium LT2, section 13 of 224 of the complete
genome.
ACCESSION AE008705 AE006468
VERSION AE008705.1 GI:16418723
KEYWORDS
SOURCE
ORGANISM
Salmonella typhimurium LT2.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 22997)
McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Flores,L., Miller,W.,
Stonking,T., Nhan,M., Waterston,R. and Wilson,R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
JOURNAL
PUBMED 11677609
REFERENCE
2 (bases 1 to 22997)
The Salmonella typhimurium Genome Sequencing Project.
AUTHORS
Direct Submission
TITLE
Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
JOURNAL
Supported by NIH grant 5U 01 AI43283
COMMENT
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
and Pedro Romero and Peter Karp at EcoCyc;
http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
ReguonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

FEATURES
Location/Qualifiers
1..22997
/organism="Salmonella typhimurium LT2"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
98..103

RBS

```

```

/gene="frr"
/EC_number="2.7.7.41"
/notes="Ortholog of E. coli CDP-diglyceride synthetase
(AAC73286.1); Blastp hit to AAC73286.1 (249 aa), 95%
identity in aa 1 - 249"
/codon_start=1
/transl_table=11
/product="CDP-diglyceride synthase"
/protein_id="AAL19186.1"
/db_xref="GI:16418727"
/translacion="MLKYLISAFVILPAVIAFLPLPPVGFALITLYVCMIAWENG
QUSGAARSORWLAIVCLGLLALMLFLPEYHNNRQPLVMSLWASLGMWVVALLL
VLFPGSAAIWRNSKTRLIFGLITVFPFWMALRAMHAYDENHYSGLATWLLYVIL
VMGADGAYMFGKLGKHLAPKVPKQWQGIIGGIATAAVISWGYGMWANNVAPV
ILLICSVAAALASVLGDLTESMFKREAGIKDSGLIPGHGGILDRIDSLTAAPVFEAC
LLLLVFRTL"
3950. .5310
/gene="yael"
/notes="STM0223"
3950. .3955
/gene="yael"
/notes="Putative RBS for yaeL; RegulonDB:STMS1H000590"
3958. .5310
/gene="yael"
/notes="Ortholog of E. coli orf, hypothetical protein
(AAC73287.1); Blastp hit to AAC73287.1 (450 aa), 88%
identity in aa 1 - 450"
/codon_start=1
/transl_table=11
/product="putative membrane-associated Zn-dependent
protease"
/protein_id="AAL19187.1"
/db_xref="GI:16418728"
/translacion="MLSILMNLAAFIAGLVITVHEFGHFWARRGCVRFERSIGF
GKALWRTDRYGEYVIALIPGLGYVYKMLDERAEVPAPELRRHAFNNKTVQORAAIT
AGPVNFIAPAYWLVFIIGVPGVPEVIGEITPNSIAOAOIAPCTELKAVDGETP
DPAVRQLVSKDQDQTTVSAPGSDQDQDITLDRHMAFEDQDQVSSIGIIPR
GQPIEPLSVKANSKAGLQAGRVKQDQDQPLQWKKFVIFVNDNPGKPLALEI
ERGSALSLTPTKTSVNGKAGFAGVPKIIPLPEYKTIQYGFSSAILBATDKT
WQMLKTVSMGLKTLGKVLNLSGIPSTIAQAGMSAEFGVIYLYMFLAITSVNLGI
INLPLPLVDGGHLLFLAIEKLGKGPVSERQVDFSYRIGSILLVLLMLALFNDFSL
"
5342. .7756
/gene="yael"
/notes="STM0224"
5342. .7756
/gene="yael"
/notes="Ortholog of E. coli orf, hypothetical protein
(AAC73288.1); Blastp hit to AAC73288.1 (810 aa), 92%
identity in aa 1 - 810"
/codon_start=1
/transl_table=11
/product="putative outer membrane antigen"
/protein_id="AAL19188.1"
/db_xref="GI:16418729"
/translacion="MAMKLLIASLFFSATVYCAEGFVVKDIHFEGLQRYAVGAALL
SMPVRTGDTVDEDSINTRALFATGNFEDVRVLDRGNTLLVQKERTIASIFSGN
KSYDDMLKQNLKASGVYRGESLDRITLSIEKLEDFYISVGYISASVAVVTPLR
NRVDLKLQEGYSAKTQIINVGNAFSTIELISHFQLRDEVPWNVGDRYKQK
LAGDLETLRSYLDRGVAFREYNIDTQVSLTPDKKGIYITVNIIEGQDYKLSGVQVSGN
LAGSABEILTKIEGELYNGTKVTKMEDDIDKLLGYGVAYPRVQSQPEINDADKT
VKLVNVDAGNRVYRKIRFEGNDTSDKSVLRMRMOMEGALGSLDLDQGERLNL
GFTETVDTQRPVSPDQVVDVYKVRNITGSFNGIGITGESGFQAGVQDQNL
GTYSYINGTKNDQTYSELSVNTFYTVDSGLGRIFYNDFQADADSLDTNKS
TDTVTLGFFINEYTPRAGLVYHNKLSNMQPIAMDRYLESMSGDSFSAADF
TYNGWTYNKLDGYPTDGRVNLTKGTIIPGSDNEYKVLSTATYVPIDNDHKWY
VLGRWGYDGLGGKEMPEYENFAGSGSTVGFQSNITGPKAYKNGAHTSWDDND
DYEDTOESGSDDDAVGGNAMAVASLEFIPPFISEKIANSVRTSEFFDMGWFDT
NWDPSAPSVDVPSDPGNIRMSAGIALQWMSPLGLVFSYAQPFKKYDGDKAQFQF
NIGKTW"

```

alignment_scores:

[illegible]

```
|||||
6991 GTCGATGGGACAACTGCTGATACCAGCAGTTTTCGCGGTGATGACTTCA 7040
574 yshisAspTyrThrTyrAsnAlaIleLeuGlyTyrPAsnTyrSerSer 590
7041 CCTTTAACTAC.....GGCTGGACCTATAACAAG 7069
591 LeuAspArgProValPheProThrGlnGly.....MetSerHisSerVa 605
7070 CTTGACCGTGGTTATTTCCGACGGAGCGGTTCGCGGCTCAATCGACGG 7119
605 lAspLeuThrValGlyPheGlyAspLysThrHisGlnLysValValTyrG 622
7120 GAAAGTACGATTCAGGTTCGACACGAACTACTACAAAGTCTCGCTGG 7169
622 lngLysnIleTyrArgProPheIleLys.....LysSerValLeuArg 636
7170 ATACCGGACCTATGTCCTGATGATAACGATCACAAGTGGGTTCG 7219
637 GlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeu.....Pr 648
7220 GGGCGTACTGCTGGGTTACGGCGATGTTTAGCGGCAAGGAATGCC 7269
648 oPheTyrGluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrA 665
7270 GTTCTATGAGAACTTCTATGCGGTGTTCCAGTACGTCGCTGGTTCC 7319
665 spGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArg 681
7320 AGTCGAATACTATCGGCCGAAAGCG.....GTCTATAAAAT 7357
682 GlyGlnGlnThr..... 686
7358 GGCCTCATCTAGTTGGGACGATAATGATTACGAGACTGTACTCA 7407
687 .....LeuGlyGluValValGlyGlyAsnAlaLeuAlat 698
7408 GGAATCAGCGTGTAATTCAGATGCGGTAGGTGTTAATGCCATGGCTG 7457
698 hrPheGlySerGluLeuLeuProLeuProPheLysGlyAsp...Tip 713
7458 TCGCACCTGGATTTATTTACCCGACGCGGTTTCATCAGCGAAAGTAT 7507
714 IleAspGlnValArgProValIlePheIleGluGlyGlnValPheAs 730
7508 GCCAACTCGTCTGATCCTCTCTCTGGGATATGGTACCGTTTGGGA 7557
730 pThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPheLysA 746
7558 TACGAATGGATCCGTCATCTCGCGCGTCAGATGTACCGGATACACG 7607
746 spProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeu 762
7608 ATCCA..... 7612
763 LeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTr 779
7613 .....GGCAACATCCGATGTCGCGGGTATCGCATTAACAAT 7650
779 pTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnL 796
7651 GATGTCCTCCATGGGCGGTGGTCTCTCTACGCCACCGCGGTTTAAAA 7700
796 yLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySerVal 812
7701 AGTAGATGAGAGACAAAGCGAGCAGTTCACAGTTTACATGGTAAACC 7750
813 Phe 813
7751 TGG 7753
```

seq_name: gb_ba:ECOTSF

```
seq_documentation_block:
LOCUS ECOTSF 91430 bp DNA linear BCT 28-MAY-1999
DEFINITION Escherichia coli genomic DNA. (4.1 - 6.1 min).
ACCESSION D83536
VERSION D83536.2 GI:4902908
KEYWORDS complete and shotgun sequencing; rpsB; tsf; smb; mukB; frr; rrf;
yaeM; cdsA; cds; yael; yzn; hlpA; skp; omp; firA; fabZ; sefA;
lpxA; lpxB; pgsB; rnhB; dnaE; polC; accA; ldcC; ldcH; yaeR; mesJ;
yaeO; yaeP; yaeQ; yaeJ; cutB; nlpe; yaeF; proS; dnpA; yaeB; rcsF;
yaeC; yaeE; abc; yaeD; yafC; yafD; yafE; yafG; dnlK; rnhA; rnh;
dasF; herA; sdrA; dnaQ; mutD; yhhI; yafH; gmbA; lcaA; trfA; yafK;
gpp; yafI; cri; phoE; ompE; proB; proA; yfjZ; yfjY; yfjX; yfjT;
yfjS; yfjQ; yfjP; dsdC; rocE; yagC; yagA; yjhH; yagB.
Escherichia coli (strain:K12) DNA.
SOURCE
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 91430)
AUTHORS Musso,R., Di Lauro,R., Rosenberg,M. and de Crombrughe,B.
TITLE Nucleotide sequence of the operator-promoter region of the
JOURNAL galactose operon of Escherichia coli
PROCEEDINGS OF THE National Academy of Sciences of the United
STATES OF AMERICA. 74 (1), 106-110 (1977)
MEDLINE 77102694
PUBMED 319453
REFERENCE 2 (bases 1 to 91430)
AUTHORS Young,R.A. and Steitz,J.A.
TITLE Complementary sequences 1700 nucleotides apart form a ribonuclease
JOURNAL III cleavage site in Escherichia coli ribosomal precursor RNA
PROCEEDINGS OF THE National Academy of Sciences of the United
STATES OF AMERICA. 75 (8), 3593-3597 (1978)
MEDLINE 79012454
PUBMED 358189
REFERENCE 3 (bases 1 to 91430)
AUTHORS Mulligan,R.C. and Berg,P.
TITLE Factors governing the expression of a bacterial gene in mammalian
JOURNAL cells
PROCEEDINGS OF THE National Academy of Sciences of the United
STATES OF AMERICA. 78 (7), 4199-4203 (1981)
MEDLINE 82037806
PUBMED 7027255
REFERENCE 4 (bases 1 to 91430)
AUTHORS Little,J.W., Mount,D.W. and Yanisch-Perron,C.R.
TITLE Purified lexA protein is a repressor of the recA and lexA genes
JOURNAL Proceedings of the National Academy of Sciences of the United
STATES OF AMERICA. 78 (7), 4199-4203 (1981)
MEDLINE 82037806
PUBMED 7027255
REFERENCE 5 (bases 1 to 91430)
AUTHORS An,G., Bendiak,D.S., Mamelak,L.A. and Friesen,J.D.
TITLE Organization and nucleotide sequence of a new ribosomal operon in
JOURNAL Escherichia coli containing the genes for ribosomal protein S2 and
elongation factor Ts
PROCEEDINGS OF THE National Academy of Sciences of the United
STATES OF AMERICA. 78 (7), 4199-4203 (1981)
MEDLINE 82059454
PUBMED 6272196
REFERENCE 6 (bases 1 to 91430)
AUTHORS Overbeek,N., Bergmans,H., van Mansfeld,F. and Lugtenberg,B.
TITLE Complete nucleotide sequence of phoE, the structural gene for the
phosphate limitation inducible outer membrane pore protein of
JOURNAL Escherichia coli K12
PROCEEDINGS OF THE National Academy of Sciences of the United
STATES OF AMERICA. 78 (7), 4199-4203 (1981)
MEDLINE 83189086
PUBMED 6341601
REFERENCE 7 (bases 1 to 91430)
AUTHORS Kanaya,S. and Crouch,R.J.
TITLE Low levels of RNase H activity in Escherichia coli FB2 rnh result
JOURNAL from a single-base change in the structural gene of RNase H
PROCEEDINGS OF THE National Academy of Sciences of the United
STATES OF AMERICA. 78 (7), 4199-4203 (1981)
MEDLINE 83185998
PUBMED 6302075
REFERENCE 8 (bases 1 to 91430)
AUTHORS Richardson,K.K., Fostel,J. and Skopek,T.R.
```



```

DIDVIBQMIFHYSLRLAIISLVGAGLGVGLVFOVLRNPLAEPTTIGVATGAQLG
ITVTTLWALPGMASQFAAQAGACVVGVLIVFGVAMGKRLSPVTLILAGVWSLYCGAI
NDLVLFTHDIOQLMSTGTOTDQMGVRLWPQLLGGVMTLLRLPLTLMLGLD
DGVARNLGLALSRLAALSLAIVISALLYNVAGIIGFGLFAPLAKMLGARRLLPR
LMSASDIERHILMSQIILMLTRVWMEVSTGVTALIGAPLLMLLPLRLRSISAPDM
KLVNRVAERQHVLAALAGVLLMLAVVVALSFGSDAHGHTWASGALLEDLMPWMP
RMAALFAGVMLAVACIIRLTQRTNPMASPEVLIGSSGAAGVVLMLFVLPVGGFWL
LPASGCAAVTLIIIMAAAGRGFSRPHMLLAGMALSTAFMLMLMQASGDPMAOV
LFWISGSTYNATDAQVWRTGIYVMIILATPLCRRLWTILPLGGDPARAVGALTPTR
IALLLLAAGLTATATWITIGLFSVGLMAPHIARMGFRRTMPHIVISALVGGLLVFA
DMGRWVLPFQIPAGLLSTFIFGAPYFIILLRQSR"
/translation="glutamine-1-semialdehyde aminotransferase"
/protein_id="AAB08584.1"
/db_xref="GI:1552732"
/translation="MSKSNLYSAARELIPGVNSPVRAFTGVGGTFLFIEKADGAYL
YVDGKAYIDYVSGWPMVLGNHHPAIRNAVIEAERGLSFGNAPTMEVYKMAQLYTEL
VPTMWRVWNSGTEATMSAIRLARGFTGRDKIIEEGCYHGADCLLVKAGSALT
GQNSPGVPADFAKYTLCTYNDLASRAAFQYPOEIACTIIEPVAGNMNCPVPLPE
FLPLRALCDEFALLIIDEVMTGFRVALAGAOQYGVVPLDTCLGIIGGGPVGAF
GRRDVMALAPTGPVYQATLSSGNPIMAAAGFACILNEVAQPGVHETLDELTRLAEG
LLEAAEAGIPLVNVHVGWGFIFFTDAESVTCQDVMACDVERFKRFFHMLDEGVY
LAPSAFEGCFMSVAHSMEDINTIDAARVFAKL"
/translation="similar to voltage-gated chloride channel protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08585.1"
/db_xref="GI:1552733"
/translation="MKTDPSLETPOAKRLRRQLRQLRLERDKTPLAILPMAVVGTT
LVGLAAVADKGVAMQNRMGALVHTADNYPLLLVAFCLSAVLAMFYFLVRYKYP
EAGSGPIEGEALEDQRPVWRWVLPVKEFGGLTGLGGVWLVGGVYVLTGGNGR
MVLIDFRLGDEARHPTLALGTAAGAAAFNAPLAGILFIEBMRPQRYTLISIKAV
FVIGVIMTYRIFNFEHALIDVGLSDAPLNTLWLYLILGIFIGFPGFINFKWILGM
QDLHRVGNNTIKVWLMGALGGLGLGFLVAPATSGGFNLIPITATAGNFSGMLV
FVFAVITITLLCFSSGAPGIFAPMLALGTVLGTAFGMVAVLELFPYHLEAGTFAIA
GMLAALASIRAPLGTIILVEMTDNYQLILPMIITGLGATLLAQTGKPLYSALIA
RTLAKQEAQSLAKSAASENT"
/translation="hypoetical protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08586.1"
/db_xref="GI:1552734"
/translation="MSDDVALPLEFTDAAANKYKSLIADENPNLRLRYIITGGGCGS
FOYGFTFDQVNBGDMTIBKQGVGLVDPMSLQYLVGGSVDYTEGLEGSRFITNPN
KSTCGCGSSFSI"
/translation="similar to Vibrio cholerae RecA protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08587.1"
/db_xref="GI:1552735"
/translation="MLVYWLDDIVGTAVFAISGVLLAGKLRMDPFGVLVGVYTVAGGG
TIDKMDLHGPVFWKDPDLDVWAVTSMLTIVLVRPQRLRKLWLPDLDVGLAVFV
GTVGNKFAAGPLIACVGVITVGGGIIIRDVLAREIPMLIRTIYATACILIGIV
HATAYTFVSVPLETSAGMGVWTLIRLAAIRWHLKLPTFALDENGR"
/translation="similar to voltage-gated chloride channel protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08588.1"
/db_xref="GI:1552736"
/translation="MAKSLFALVALSFLAPLWLNAPRVITLSPANTELAFAAGITTP
VGVSSYSDPPQAKTEQVSTWQGMNERIALVKPDVLTAMRGNAEROVDOQLASLGI
KVMVDPTTSEIQANALQALAPSPQPKRAEQAAQSILLOYAKLQAKYADPKRRVFL
QVGNPATTSGRESIQNVLEVGGENIFRDSRVPWPVQVSRQVLAARSPOAIVITGGP
DQPKTKQWGEOLKIPVPLTSDWFERASPRILAAQAQCLNALSQVD"
/translation="similar to purine nucleoside phosphorylase (deod)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08589.1"
/db_xref="GI:1552737"
/translation="MKIIGIIGAMEEVEVTLTRDKIENROTISLGGCEIYTGOLMGTEVA
LLKSGIGKVAALGATLLIEHCKPDVLTNGSAGGLAPTLKVGDIIVSDEARYHDADV
TAFGYEGQLPGCPAGFKADDKLIAAEACIAELNNAVRLGLVSDAFNGSVGLAK
IRNFPOAIAVEMETAIAIHVCHNFVNFVVRVRAISDVADQQSHLSFDEFELAVAAKQS
SLMVSLSVQKLAHG"
/translation="deoxyuanosine triphosphate"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08590.1"
/db_xref="GI:1552738"
/translation="MAQIDFRKKNHRRYRSQGVKTHEILRIFESDRGRILNSPA
IRLQOKTVFPLERNAVTRTLTHSMEOVQVGRYIAKEILSLKELKLELGLDEL
TGPFESIVMSCLMHDIGNPPFHFGEAIIINDFWRQRLHPEDAESQPLTDDRCVSAAL
RLRDGEPLNELRRKIQDLCHPEGNAQGIRLVHTLRMNLWTAQVGGILKYTRPAAW
RGTPETHVLMKKPGYILSEAYIARLRELNALYSRPLTWIMEAADDISYCVAD
LEDAVEKRIFTVEQLYHHLHEAMGQHEKGSLSFLVVENAWEKSRNSLSRSTEDQFFM

```

alignment_scores:

	Quality	Length	848
	Ratio	Gaps	15
	Percent	Similarity	64.976
	Percent	Identity	33.844
alignment_block:			
US-09-701-711-2 x ECU70214 ..			
Align seg 1/1 to: ECU70214 from: 1 to: 123171			
10	GlnValSerAlaMetThrMetAlaValMetMetValMetSer.....	23	
28991	CGCATAATAACGATGGCGATGAAAAGTTGCTCATAGCGTCTGCTGTT	29040	
24ThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleT	37	
29041	TAGCAGCGCCACCGTATACGGTGTGAAGGGTTCGTAGTGAAGATATTC	29090	
37	hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu	53	
29091	ATTTCGAAGGCCTTCAGCGTTCGCCGTTGTGGCGCCCTCTCAGTAGT	29140	
54	ProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVa	70	
29141	CCGGTGCACAGCGACACGGTTAATGATGAAGATATCAGTAAATACCAT	29190	
70	llyAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG	87	
29191	TCGGCTCTGTTGTACCGGCACTTTGAGGATGTCGCGCTCTCGTG	29240	
87	lnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGlu	103	
29241	ATGGTGATACCTTCTGTTTTCAGGTAAAGACGTCGCCACCATTCACG	29290	
104	IleAsnPheGluGlyAsnArgLeuIleProlysGlyLeuGlnGluGlu	120	
29291	ATTACTTCTCGGTACAAATCGGTGAAGATGACATGCTGTAAGCAAAA	29340	


```

111
31096 CTGTGTAATCGGATGCTGTAGCGGTACGCCATCGCGGTGCCA 31145
      :::::::::::|||||
700  lysergluLeuLeuProLeuProPheLysGlyAsp...TrpIleAsp 715
      :: |||::||| ||| |||::||| ::||| ::|
31146 GCTCGAGTTCATCACCGCGCGCGTATTATAGCGATAAGTATGCTAAC 31195
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
716  GlnValArgProValIlePheIleGluGlyGlyGlnValPheAspThrH 732
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
31196 TCGGTTCTGACTCTCTCTCTGGGATATGGGTACCGTTTGGGATACAA 31245
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
732  rGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnA 749
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
31246 CTGGGATTCAGCCATATCTCGGATATCGCGACTATAGTATGCCA... 31291
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
749  laThrAlaGluGlnAsnAlaLysAlaAAsnArgProLeuLeuThrGln 765
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
31291 ..... 31291
766 AspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTyrThrPr 782
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
31292 ...AGCAATATCGTATCTCGCGGTATCGCATATACATGGATGCCCC 31338
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
782  oIleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnA 799
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
31339 ATTGGGGCGGTGCTGCTCTCTACGCCCGCGCTTCAAAAAGTACGATG 31388
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
799  snAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
31389 GAGACAAAGCAGAACAGATTCCAGTTTAACATCGGTAAACCTGG 31432
      ::|||::||| ::||| ::||| ::||| ::||| ::|||
seq_name: gb_ba:AP002550

seq_documentation_block:
LOCUS AP002550 281530 bp DNA linear BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 1/20.
ACCESSION AP002550 BA000007
VERSION AP002550.1 GI:13359456
KEYWORDS .
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
AUTHORS 1 (sites)
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20198780
2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain M61655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182

4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 281530)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.
COMMENT Location/Qualifiers
FEATURES
source
1. 281530
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
190..273
/gene="Ecs0001"
190..273
/gene="Ecs0001"
/note="similar to THRL_ECOLI g11786182 percent similarity
100 in 21 aa, but has 6 additional residues (Conserved in
E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="thr operon leader peptide"
/protein_id="BAB33424.1"
/db_xref="GI:13359457"
/translation="MKRISTITITITITITITITITITITGNGAG"
354..2816
/gene="Ecs0002"
354..2816
/gene="Ecs0002"
/note="similar to THRA_ECOLI g11786183 percent identity
99 in 820 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="aspartokinase I-homoserine dehydrogenase I"
/db_xref="GI:13359458"
/translation="MRVLKFGGTSVANAERFLRVADILESNAEQGVATVLSAPAKIT
NHLVAMIEKTSIGDAPLNSIDAERIFAELLTGLAAQGFPLAQLKTFVDFQFAIK
HVLHGISILGQCPDSINAALICRGKMSIATMAGLEARGHNTVIDPVKLLAVGHY
LESTVDIAESTERIAASRIPADHVMYAGTAGNEKGYLVLRNGSDYSAVLAACL
RADCCIEITWDVGVYTCDPQVPRDARLLKSMYSIQEAMLSYFGAKVLHPRTTPTAQF
QIPCLIKNTGAPFTLIGASRDEDELFPVKGISLNNMAMFVSVPQMGKMGVMAAR
TERLAIISVVGDMRTLIGISAKFFAALARANIVATAQSSERSISVYVNDATT
GVRYTHOMLENTDOVIEVIGCVGGGALLPOLKROOSLKNKHIDLRVCGVANSKA
LLTNVHGLNLENQWQELAQAEPPNGLRLRLVKEYHLLNPVVDCTSSOAVDOYAD
FLREGFHVVTNNKANTSSMDYIHLLRHAEEKRRKFLYDYNVAGLVPVLENQLNLLN
AGDELKFSGLISLSLYIFGKLDEGMSFSEATTLAREMGYTEPPDRDLSQMDVARK
LLILARETGRELEADIETPVLPAEFNAEGDVAAFMANLSQDLDFARVAKARDEG
KVLRVGNIDEGVCRVKIAEVDGNDPLFKVKNGENALAFYSHTYQPLPLVLRGYAG
NDVTAAGVFADLLRTLSNKLGV"
2818..3750
/gene="Ecs0003"
2818..3750
/gene="Ecs0003"
/note="similar to THRB_ECOLI g11786184 percent identity
99 in 310 aa (Conserved in E.coli K-12)"
```

```

/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="homoserine kinase"
/protein_id="BAB33426.1"
/db_xref="GI:13359459"
/translation="MKVYAPASNMVSQVDFYLAAGAVTPVDGALLGDVVTVESAEFF
SNNLGRFADKLPSEPRENIVVQWERFQCELGKQIPVAMTLEKNMPPIGSLGSSACS
VVAALMANHEGCKPLNDRLLALMGELEGRISGTHYDNVAPCFRLGGWQMTEENDI
ISQOVPGEDEWLVLAYPGIKVYSTAEARAILPAQYPRQDCIAHGRHLAIFHACYSRQ
PELAALMKDVIAEYPRRELLPFQRAQVAEIGAIVASGISGSGPTLALCDKPDPA
QRVADWLGKNIQONQEGFVHICRLDTAGARVLEN"
3751..5037
/gene="ECs0004"
CDS
3751..5037
/gene="ECs0004"
/notes="similar to THRC_ECOLI gi11786185 percent identity
99 in 428 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="threonine synthase"
/protein_id="BAB33427.1"
/db_xref="GI:13359460"
/translation="MKLYNLKDHNEQVSQAQVTOGLGRKNOGLFFPHDLPERSLTEID
EMLKDFVTRSAKILSAFTIGDEIPOEILERRVRAAFAPFAPVANVEDVGCLELPHGP
TLAKDFGGRNAOMLTHTAGDKPVTILTATSGDCAAAVAHAFYGLPNVKVLYPRG
KTSPOEKLEFCTLGNIEVTAIDGDEDAQALYKQAFDDEELKVALGLNSANISLR
LQAQCTFEAVALQPEARNQLVVSQNFNGDLTAGLAKSLGLPLVPRKFAANTNVN
DIVRFLDHQSPKATQATLSNMDVSPQNNMVRVEELFRKKIWLKELGYAAVDDDE
TTQOTMERKELGYSSEPHAAVAYRALRDLNPGEYGLFLTAHPAKFKESVEAILGE
TLDLPKELAEARADLPLLSHNLPAFALRKLMMNHQ"
5251..5547
/gene="ECs0005"
CDS
5251..5547
/gene="ECs0005"
/notes="similar to B0005_ECOLI gi11786186 percent identity
95 in 98 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB33428.1"
/db_xref="GI:13359461"
/translation="MKMQSIVLALSILVLPVMAQAAEITLVPSVKLQIGDRDNRGY
YNDGHWDRDHGKWOHYEWRGNRHPGPPPPRHHKKAHHDDHGGHGGKRRH"
complement(5700..6476)
/gene="ECs0006"
CDS
complement(5700..6476)
/gene="ECs0006"
/notes="similar to YAAA_ECOLI gi11786187 percent identity
99 in 258 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="hypothetical protein"
/protein_id="BAB33429.1"
/db_xref="GI:13359462"
/translation="MLILISPAKTIQSPLTTRYTLPELLDMSQOLIHAEKLTTPP
QISTLMRISDKLAGNARFHDWQPDFTPENARQAALAFKGDVYITGLQAEITFSEDDFD
FAQOHLRSLSGYLVRLPDLMPQYRLENGARLENARKDLYQFWGDIITNKLNEALA
AQGDNVVINSLEDFYKSVKPKLNAETIKPVFLDEKNGKFKIISFYAKKARGLMSRF
IITENRITKPEQLTGNSGEGYFFDEASSNGELVFKRYEQR"
complement(6546..7976)
/gene="ECs0007"
CDS
complement(6546..7976)
/gene="ECs0007"
/notes="similar to YAAJ_ECOLI gi11786188 percent identity
99 in 476 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="putative inner membrane transport protein"

```

```

/protein_id="BAB33430.1"
/db_xref="GI:13359463"
/translation="MPDFSFINSVLWGSVMYIYLLFGAGCWFTFRTGFVQFVRIROPFG
KSLKNSIHQPQGLTSEFQSLCTSLAARVSGNLAGVALAITAGGPGAVFWWVAAFIG
MATSFAECSLAQLYKERDVGQFRGGPANTMARGLMRGMGVLFVAFLLIAYGIIFSG
VQANAVARALSFDFPPLVTGIIILAVFALLAIIIRGLHGVARLMQGFVPLMAIITWILT
SLVICVINTGQLPHVTSIFESAFGWEAAGAAAGTYSQAITNGFQSRMFSNEAGMG
STPNAAAAASWPPHAAQGIQVMIGIFIDTLVITASAMLILLAGNGTTPMPLEGIQ
LLOKAMRVLMGSGWAEFVTLVILFAESSIVANVIAENNLFFLRLLNPKAIWCLRTIC
TRATVYGGTLLSLPLMWQLADIIMACMAITNLAILLLSPVHTIASDYLQRKRLGVR
PVFDPLRYPDIDQLSRDAWDDVSQE"
8255..9208
/gene="ECs0008"
CDS
8255..9208
/gene="ECs0008"

alignment_scores:
  Quality: 1212.00      Length: 848
  Ratio: 2.200         Gaps: 15
  Percent Similarity: 64.976  Percent Identity: 33.844

alignment_block:
US-09-701-711-2 x AP002550 ..

Align seg 1/1 to: AP002550 from: 1 to: 281530

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSer..... 23
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201260 GGCATAATAAGATGGCGGATGAAAAGTTGCTCATAGGCTGCCTGCTGT 201309

24 .....ThrHisAlaGlnAlaAlaAaspPheMetAlaAsnAspIle 37
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201310 TAGCAGCCACCCTGATACGGTGTGAAGGGTTCGTAGTGAAGATATTC 201359

37 hrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeu 53
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201360 ATTTCGAAGGCTTCAGCGTGTGCGGTTGGTGGCGGCTCTCCTAGTAG 201409

54 ProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAaspGlyVa 70
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201410 CCGTGGCCACAGCGCACCGGTANTGATGAAGATACAGTAATACCAT 201459

70 LysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisG 87
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201460 TCGCGCTCTGTTGCTACCGCAACTTTGAGGATGTCGCGCTCCTCGTG 201509

87 lnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaGlu 103
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201510 ATGCTGATACCTCTCTGTTCCAGGTAAAGAACGTCGACCATTCGCAGC 201559

104 IleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201560 ATTACTTCTCCGGTAACAAATCGGTGAAGATGACATGCTGAAGCAAAA 201609

120 yLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrV 137
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201610 CCTCGAGGCTCTCGTGTGCGGTGGGCGAATCCCTCGATCGCACCAACA 201659

137 alGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 153
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201660 TTGCGCATATCGAAGAGTGTGGAAGACTTCTACTACAGCGTCGGTAAA 201709

154 TyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVa 170
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201710 TATAGCGCCACGGTTAAGCTGTGTCGCCGCTGCCCGCAACCGGTGT 201759

170 lLysLeuAspMetThrPheAlaGluGlyProAlaArgValValAspI 187
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201760 TGACCTAAACTGGTGTTCAGGAAGGTGTGTGCTAGCTGAATCCAGCAA 201809

187 leAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 203
: : : : : : : : : : : : : : : : : : : : : : : : : : :
201810 TTAACATTGTTGTTAAACCATGCTTTCACCACCGCAGCAACTGATCTCAT 201859

```

204 LeuAlaIleLysAsp.....AsnLysIleAsnProLeuSe 215
::: ::::: ||| :::::
201860 TTCACACTGGCGTACCAAGTCCGCTGGTGGAACTGGTA..... 201898
215 rLysAlaAspArg...TyrThrGlnGluLysLeuValThrSerLeuGluA 231
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
201899GGCGATCGTAATACCAAGAACAGAACTGGCGGGCGCTTGAAA 201944
231 snLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLys 247
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
201945 CCCTGGCGACGCTACTCTCGACCGCGGTATGCCGTTTCAACATAGAC 201994
248 AsnAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluI 264
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
201995 TCTACCCAGGTGACGTGACGCCAGATAAAAAAGGTATTTACGTCCGGT 202044
264 eSerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuG 281
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202045 GAACATCACCGAAGCGGATCAGTACAAAGCTTCTGGCGTTGAAGTGAGCG 202094
281 lYasnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLysPheLys 297
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202095 GCAACCTTGGCGGCATCCGCTGCAATTTAGCAGCTGACTAAGATCGAG 202144
298 AlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnI 314
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202145 CCGGCGAGCTGTATACGGCACCAAGTGACCAAAATGGAAGATGACAT 202194
314 eSerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProV 331
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202195 TAAAAAGCTTCTCGGTGGCTATGGTTATGGCTATCCGCGCTACAGTCGA 202244
331 alThrArgIleAsnAspGluSerArgThrValaspValGluTyrTyrIle 347
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202245 TGCTGAAATTAACGATGCCGACAAACCGTTAAATACGGCTGAAGCTT 202294
348 AspProValHisProValTyrValArgIleAsnPheThrGlyAsnPh 364
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202295 GATCGGGTAACCGTCTCTACGTGCGTAAGATCCGTTTGAAGGTAAACA 202344
364 eLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyA 381
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202345 TACCTGAAAGATCCGCTCTCGCTCGCAAAATCGCTCAGATGAAGGTG 202394
381 lateuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArg 397
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202395 CATGGCTGGGAAGCGATCTGTCGATCAGGTAAAGGAGCGTCTGAATCGT 202444
398 ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202445 CTGGGCTCTTTGAAACTGTGATACCGATACCCAAACGTTTCCGGGTAG 202494
414 rProAspGlnValAspValAsnPheValValGluGluGlnProSerGlyS 431
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202495 CCGGACAGGTAGTAGTCTAGCTACAGGTGAAAGACGTAAACACTGGTA 202544
431 eSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGln 447
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202545 GCTTCAACTTCGGGATTTGGTTACGGTACTGAAAGTGGCGTGAGTCCAG 202594
448 PheAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAl 464
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202595 GCTGGCGTGCACAGGATAACTGTTAGGTAGTACAGGTTATGCTGTTGGTAT 202644
464 aSerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrA 481
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202645 CAACGGGACCAAAACGATTAACGACCATGCTGGAACCTGCGGTAAACA 202694
481 snProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyr 497
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202695 ACCCGTACTTCAACCGTAGATGGCGTAAGCCTCGGTGGTGGTCTCTCTAT 202744

498 ArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSe 514
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202745 AATGACTTCCAGGCAGATGACGCCGACCTGTCGCACTATACCAACAGAG 202794
514 rTyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgI 531
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202795 TTATGTTACAGAGCTGACGTGGCTTCCGATTAACGAATATAACATCCG 202844
531 leSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPhe 547
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202845 TGGCTGCAGGTCTGGGTATGTTACATAAATCCCTG..... 202879
548 MetGlyIleSerAsnValLysGlnLeuMetAla..... 558
202880TCCAACATGCAGCCTCAGTTCCGATGTGGCTTATCTGTGTA 202920
559 AspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysH 575
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202921 CTCTATGGGTGAACATCCGAGCACCTCTGATCAGGATAACAGTTCAAAA 202970
575 isAspTyrThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeu 591
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202971 CGGACGACTTTCAGCTTCAAC.....TATGTTGGACCTATAACAGCTT 203014
592 AspArgProValPheProThrGlnGly.....MetSerHisSerValas 606
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203015 GACCGTGGTTACTTCCGACAGATGTTTCACGTGTCAACCTGACCCGGTAA 203064
606 pLeuThrValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnG 623
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203065 AGTGACCATTCCTGGATCGGATACGAATACTACAAAGTACGTTAGACA 203114
623 lYasnIleTyrArgProPhe.....IleLysLysSerValLeuArgGly 637
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203115 CGGCGACTTATGTCGCGATCGATCAGCATCAAAATGGTTGTTCTGGG 203164
638 TyrAlaLysLeuGlyTyrGlyAsnAsnLeu.....ProPh 649
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203165 CGTACCCGCTGGGTTATGGTATGGTGGTGGCGGCAAGAGATGCCGT 203214
649 eTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspG 666
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203215 CTACGAGAACTTCTATGCGGTGTTCCAGCACCGTGGCTGCCAGT 203264
666 lnSerSerLeuGlyProArgSer.....GlnAla..... 675
203265 CCAATAAATGGTCCGAAAGCAGTTTACTTCCGCGCATCAGGCCAGTAAT 203314
676TyrLeuThrAlaArgArgGlyGlnGlnThrTh 686
203315 TATGATCCGGACTATGATTACGATGTCGACTCAGGACGGCGGCAAGA 203364
686 rLeu.....GlyGluValValGlyGlyAsnAlaLeuAlaThrPheG 700
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203365 CCTGTGTAATCGGATGATGCTGTAGGCGGTAAAGCCATGCGGTTGCCA 203414
700 lYserGluLeuLeuProLeuProPheLysGlyAsp...TrpIleasp 715
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203415 GCCTCGAGTTTCATCACCAGCGCGGTTATTAGCGATAAGTATGCTAAC 203464
716 GlnValArgProValIlePheIleGluGlyGlnValPheAspThrTh 732
||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203465 TCGGTTCTGCTACTCTCTCTCGGATGATGGTACCGTTGGGATACAAA 203514
732 rGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnA 749
::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203515 CTGGGATCCAGCAATATCTCGGATATCCGACTATATAGTATCCA.... 203560
749 laThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGln 765
203560 203560
766 AspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrPr 782

[illegible][illegible]

seq_name: qb_ba:AE004297

seq_documentation_block:

seq_documentation_block: BCT 31-JUL-2000
 LOCUS AE004297 10901 bp DNA linear
 DEFINITION Vibrio cholerae chromosome I, section 205 of the complete
 chromosome.
 ACCESSION AE004297 AE003852

AUTHORS

Simpson, A. J. G., Reinach, F. C., Arruda, P., Abreu, F. A., Acencio, M., Alvarenga, R., Alves, L. M. C., Araya, J. E., Baia, G. S., Baptista, C. S., Barros, M. H., Bonaccorsi, E. D., Bordin, S., Bove, J. M., Briones, M. R. S., Bueno, M. R. P., Camargo, A. A., Camargo, L. E. A., Carriaro, D. M., Carier, H., Colauto, N. B., Colombo, C., Costa, F. F., Costa, M. C. R., Costa-Neto, C. M., Coutinho, L. L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorri, H., Facincani, A. P., Ferreira, A. J. S., Ferreira, V. C. A., Ferro, J. A., Fraga, J. S., Franca, S. C., Franco, M. C., Frohme, M., Furlan, L. R., Garnier, M., Goldman, G. H., Goldman, M. H. S., Gomes, S. L., Gruber, A., Ho, P. L., Hoheisel, J. D., Junqueira, M. L., Kemper, E. L., Kitajima, J. P., Krieger, J. E., Kuramae, E. E., Laigret, F., Lampaes, M. R., Leite, L. C. C., Lemos, E. G. M., Lemos, A. M. F. B., Lopes, S. A., Lopes, C. R., Machado, J. A., Machado, M. A., Madeira, A. M. B. N., Madeira, H. M. F., Marino, C. L., Marques, M. V., Martins, E. A. L., Martins, E. M. F., Matsukuma, A. Y., Menck, C. F. M., Miracca, E. C., Miyaki, C. Y., Monteiro-Vitorello, C. B., Moon, D. H., Nagai, M. A., Nascimento, A. L. T. O., Netto, L. E. S., Nhani Jr., A., Nobrega, F. G., Nunes, L. R., Oliveira, M. A., de Oliveira, M. C., de Oliveira, R. C., Palmieri, D. A., Paris, A., Peixoto, B. R., Pereira, G. A. G., Pereira Jr., H. A., Pesquero, J. B., Quaggio, R. B., Roberto, P. G., Rodrigues, V., de M. Rosa, A. J., de Rosa Jr., V. E., de Sa, R. G., Santelli, R. V., Sawasaki, H. E., da Silva, A. C. R., da Silva, F. R., da Silva, A. M., Silva Jr., W. A., da Silveira, J. F., Silvestri, M. L. Z., Siqueira, W. J., de Souza, A. A., de Souza, A. P., Terenzi, M. F., Truffi, D., Tsai, S. M., Tsubako, M. H., Vallada, H., Van Sluys, M. A., Verjovski-Almeida, S., Vettore, A. L., Zago, M. A., Zatz, M., Meidanis, J. and Setubal, J. C.

Direct Submission

Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

FEATURES

Location/Qualifiers

1..11711

/organism="Xylella fastidiosa 9a5c"

/db_xref="taxon:160492"

/clone="9a5c"

complement(84..1424)

/gene="XF1037"

complement(84..1424)

/gene="XF1037"

/note="similar to SPIP28183 (percent identity: 65 %/query alignment coverage: 100.0 %/subject alignment coverage: 96.3 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="adenosylhomocysteinase"

/protein_id="AAF83847.1"

/db_xref="GI:9105979"

/translation="MPGLMSIRRYASKQPLKGVYVTCGLHMTQTAVLIETLKDIGA DWRWACNIFSTQDHAATAATSGTPFAWKGTELEEDCTLQALFTTISDGLTGP ELIVDGDGDTLLHKGYELNGSTWVDESDLEQVIRKLLKRIAREPGYTRVY NDWKGVSRETTGVRLYQIAATGRLVPAINVNSVTKSKFDNLGCRSLADGLKR AMDVMLAGKAVCGYGVGKGSASHSLRAYCARVITFDIPDIALQAAEGFEVTTVE DTLGADIYVTTGNKQVIRIHEHTAMKQDVIVCNIGHFNEIQIDALNLAGVKQIN IKPQVDFILPNGLNTFLLAEGRLVNLGCTGPHFSVMSNSFANGTILAQIDWQNDV YEKNYRLPKLDEVARLHLEKIGVKLTTLTANQAAAYLISVEGPKPEHYR"

2470..3981

/gene="XF1038"

2470..3981

/gene="XF1038"

/note="similar to GI12808752 (percent identity: 23 %/query alignment coverage: 104.4 %/subject alignment coverage: 98.7 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

/codon_start=1

/transl_table=11

/product="peptide synthase"

/protein_id="AAF83848.1"

/db_xref="GI:9105980"

/translation="MVHCVCLVYSKRCSECIHCRVAGLVLPVCMQTSVCLHGD DLDCLYVMSIVAPSSGSVASLPMGVSLRPLVWVGKPIITIGFLAORVGLAAOL PSGRVYNLCEDRFLFLAFYACALRGQVSLPSSRAPGVVAEIIYAAPDSYCLGDRE ILSLESLRYWHLLPEIFELDGAIPSLSDAVVAIGTSGTGVTISNLKATVISTE

gene

CDS

QDWAFHSLWSQVTVVATVPQPHYGMFEFVLLPLLPVAVVYAGRPFFPDDVAQAL AEYAPRLVTVPHLRLVQSGLRFPPEVAGIVSATAPLTHDLAVAAEACFGCCLREY FGSECTFAVTRTVYEQATRDGVSEIENADVTWYARHLPAAPMAIADFDVYEDG RFEVLGQADLELLEIAGKRSIGELNRCLLAIPGVGDALYQMPLEADQTVGRIAVVV APTLHEVHIQVKIRTKIDPLFLPRLRRKVDALPNNETGKLPREALALLADVC"

gene

CDS

4249..4413
/gene="XF1039"
4249..4413
/gene="XF1039"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

4430..4591
/gene="XF1040"
4430..4591
/gene="XF1040"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

4591..4752
/gene="XF1041"
4591..4752
/gene="XF1041"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

4752..4913
/gene="XF1042"
4752..4913
/gene="XF1042"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

4913..5074
/gene="XF1043"
4913..5074
/gene="XF1043"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

5074..5235
/gene="XF1044"
5074..5235
/gene="XF1044"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

5235..5396
/gene="XF1045"
5235..5396
/gene="XF1045"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

5396..5557
/gene="XF1046"
5396..5557
/gene="XF1046"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

5557..5718
/gene="XF1047"
5557..5718
/gene="XF1047"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

5718..5879
/gene="XF1048"
5718..5879
/gene="XF1048"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

5879..6040
/gene="XF1049"
5879..6040
/gene="XF1049"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

gene

CDS

6040..6201
/gene="XF1050"
6040..6201
/gene="XF1050"

/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"

```
/note="similar to SP|P10440 (percent identity: 44 %/query alignment coverage: 98.5 %/subject alignment coverage: 98.9 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine acyltransferase"
/protein_id="AAF83853.1"
/db_xref="GI:9105985"
/translation="MKKHSILHPTAVIAPSATLAPDVOIGACTLIGNDVQIDTGTII
GSUCTHGPTRIGNNRFIOGAIGEPQDKFAGERTELLIGNNTIRFVTLNRT
GGGGVTSIGNDWILAYTHIAHDCVGHHCFSNNASLGHVTVGDWVIFSGFSGAH
QFCRIGRYAPIGMTLGDVPPFTLIGSDTLGRGINSEGLKRRNFTPERITAIKR
AYRLYVAGLPLAEAKQOVAEQAKDNDIKELLQFIETARPLLR"
/gene="7588.. .8070)
/complement(7588.. .8070)
/gene="XF1044"
/complement(7588.. .8070)
/gene="XF1044"
/note="similar to GI|1718488 (percent identity: 49 %/query alignment coverage: 91.2 %/subject alignment coverage: 98.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="(3r)-hydroxymyristoyl ACP dehydratase"
/protein_id="AAF83854.1"
```

alignment_scores:

Quality:	1172.50	Length:	823
Ratio:	2.188	Gaps:	13
Percent Similarity:	65.128	Percent Identity:	33.779

alignment_block:

US-09-701-711-2 x AE003941/rev ..

Align seg 1/1 to reverse of: AE003941 from: 1 to: 11711

```
14 MetThrMetAlaValMetMetValMetSerThrHisAlaGlnAlaAlaAs 30
:::|||||::: ::: ::: ::: ::: ::: ::: ::: ::: :::
11599 CTAACCATTTGCTGCCAATTTACGTTGCTGCTTCACACAGCGCGCGA 11550
30 p...PheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrI 46
: |||||::: |||||::: ::: |||||::: ::: |||||:::
11549 ATCCTTCGTTGCAACGATATACGGCTGACGCGCTACAGCGCATCGCAT 11500
46 leGluSerLeuGlnSerValLeuProPheArgLeuGlyClnValValSer 62
::: ::: ::: ::: |||||::: ::: |||||::: ::: |||||:::
11499 CGGGCAGCGTCTTTACTTACCTTACCTGTTTAAACGGGGAGACACTGTGGAT 11450
63 GluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPh 79
::: ::: ::: |||||::: ::: |||||::: ::: |||||:::
11449 GACGCTAAGTTGCTGATCAATCCGTCGGCTGATCCGCACTGGTTTTT 11400
79 eSerAspValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValT 96
|||||::: ::: |||||::: ::: |||||::: ::: |||||:::
11399 TGAACACGTTGCGAGACCGCAAGGTAATATATTGGTCGCAAGTCA 11350
96 hrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIle 112
|||||::: |||||::: ::: ::: ::: ::: |||||::: |||||:::
11349 AAGAACGCGCGCAATTAATAGCTGACCATCACAGGAATAAGGACATC 11300
113 ProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValG 129
::: |||||::: |||||::: ::: |||||::: ::: |||||:::
11299 AAGTCGGAAGAACTCTCTCAAGGCGCTATCTGAGATCGGACTGAGCGAAG 11250
129 yGlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrA 146
| ::: ::: ::: |||||::: ::: |||||::: ::: |||||:::
11249 AGGTACCTTCGACCGCTCAGCGCTGACCGCTTATACCAAGAGCTGAAC 11200
146 snGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGln 162
::: |||||::: ::: |||||::: ::: |||||::: ::: |||||:::
11199 GCCAGTACAACACCGCGGTAATAACACGTTTCAGATGACAACAACGACC 11150
```

```
163 ThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluG 179
||| ||||| |||||::: ::: ::: ::: |||||
11149 ACTCCATTGGATCGTAACCGTGTGGATGTAACGATCGCGATTAAAGAAG 11100
179 yLysProAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheS 196
||| |||||::: |||||::: |||||::: |||||::: |||||
11099 CAAGGCTGCGCAAAATCCGCGACATCAATCTGATCGCGACTGAAAAATTC 11050
196 eAspAlaAspLeuIleAspValLeuAlaIleLysAsp...AsnLysIle 211
::: ::: |||||::: |||||::: |||||::: |||||::: |||||
11049 ATACAAAGATGTCAATGAGTGCCTCGGAATCCAAAGAACACAACTGGCT 11000
212 AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrS 228
::: ::: |||||::: |||||::: |||||::: |||||::: |||||
10999 TCCTGTGTACCGCGCTGACGACCAATACTCTAAAGAAAAGCTGTCTGGAGA 10950
228 rLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheG 245
: |||||::: |||||::: |||||::: |||||::: |||||::: |||||
10949 CCTTGAGAAGCTGAACCTCTGTGTACTTGGATCGTGTGCTCGATTTC 10900
245 luIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePhe 261
: |||||::: |||||::: |||||::: |||||::: |||||::: |||||
10899 ATATTGACTCAACTCAAGTGTGATCAGCCCCGAAAGAACACAACTGTT 10850
262 ValGluIleSerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrG 278
::: ::: |||||::: |||||::: |||||::: |||||::: |||||
10849 ATTACTGCGCGGCTCACAGAGCGGATCAATACAAGATCTCTCCATCAA 10800
278 nPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeu 295
: |||||::: ::: |||||::: |||||::: |||||::: |||||
10799 GGTAAACAGGCAACACCGTATTACTCTCAGGAAAAAATCGAAAAGCTAGTCA 10750
295 yspPheLysAlaGluGlyPheSerGlnAlaMetLeuGluGlnThrThr 311
|||||::: ::: |||||::: |||||::: |||||::: |||||::: |||||
10749 TCCCCAAGACAGGTGACATTTCTCGCGGCTTCTGCTGGAATACAGCTCG 10700
312 AsnAsnIleSerThrLysPheGlyAspGlyTyrTyrTyrAlaGlnI 328
||| ::: |||||::: |||||::: |||||::: |||||::: |||||
10699 GCAGCGCATCAATAATACGTCGAGCAATATCGGCTACGCATTTTCCAAGGT 10650
328 eArgProValThrArgIleAsnAspGluSerArgThrValAspValGluT 345
::: |||||::: |||||::: |||||::: |||||::: |||||
10649 CAACCAATCCCAACACCAACCGCTGACCGTACCGCTCGCGCTCAACT 10600
345 yTyrIleAspProValHisProValTyrValArgArgIleAsnPheThr 361
::: ::: |||||::: |||||::: |||||::: |||||::: |||||
10599 TGCAGCTTATCCCTGGCGCTCGGCTCAGCTGCGTCAGATCTCTGTTCAA 10550
362 GlyAsnPheLysThrGlnAspGluValLeuArgArgGluMetArgGlnLe 378
||| ::: |||||::: |||||::: |||||::: |||||::: |||||
10549 GGTAAATACCGCACCTCCGATGAAGTACTACGTCGTGAAATCGTCAGTT 10500
378 uGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgL 395
: |||||::: |||||::: |||||::: |||||::: |||||::: |||||
10499 CGAAAAACAGTTGTTACTCGCAACGACGCGATTGACCGCTCCAGATACGCC 10450
395 euMetArgThrGlyPhePheLysHisValThrValAspThrArgProVal 411
|||::: |||||::: |||||::: |||||::: |||||::: |||||
10449 TACAACGCTTGGCTACTTTCGAAGCGTGGACGTAGAGACACACAGTT 10400
412 ProAsnSerProAspGlnValAspValAsnPheValValGluGlnPr 428
|||::: |||||::: |||||::: |||||::: |||||::: |||||
10399 CCAGGAAGCAACGACCAAGTGGATATCGTCTATACCGTCAAGAGACTAC 10350
428 oSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSerGlyValT 445
|||||::: ::: |||||::: |||||::: |||||::: |||||
10349 CTCGGGTAGTCTCCAGGTGCGCTTGGTTACTCGAAAAACCTATGGTGTCA 10300
445 hrPhePhePheValSerGlnAsnAsnPheMetGlyThrGlyLysHis 461
||| ::: |||||::: |||||::: |||||::: |||||::: |||||
10299 CCACCTCAGTACAGTTATTCGAGAACAAATTTCTTGTGTAGTGCACCCGC 10250
462 ValAsnAlaSerPheSerArgSerGluThrArgGluValTyrSerLeuG 478
```

```
|||||..... |||||||..... :..... |||||||.....
10249 GTGTCGTAGATGCTTCACGCAGACAGATACCAAGATCGCTATTGTTCTC 10200
478 yMetThrAsnProTyrPheThrValAsnGlyValSerGlnSerLeuSerG 495
: |||||||..... |||||||..... : : : : :
10199 CTACACCAACCGTTTTTCACCGCAATGGGTCTCTCAGTACGATATAACT 10150
495 lyTyrTyrArgLysThrLysTyrAspAsnLysAsnLysAsnLysAsnLysVal 511
|||||..... |||||||..... |||||||..... |||||||.....
10149 TGGCCTACCAAAATACACTACTCAGACTTCACGCTGCGCAATACAAAC 10100
512 LeuAspSerTyrGlySerLeuSerTyrGlyTyrProIleAspGluAs 528
: : : : : |||||||..... : : : : : |||||||.....
10099 AGCAAGCGTATGTCAGGACAAACCATCTTTGGAATCCCGATTACAGAA 10050
528 nGlnArgLysSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyG 545
: : : : : |||||||..... : : : : : |||||||.....
10049 CGACACTGCTCTTGGGTGTCATGGCGCGAGATAGCAACCAATC.....A 10006
545 lyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGly 561
||| ||| ||| ||| ||| : : : : : ||| : : : : :
10005 CGACCTTCCAGGT...TCCAGCGCAAAAGCAATTTATCGAC..... 9968
562 LysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTyrTh 578
: : : : : ||| : : : : : : : : : : :
9967 .....TACATCGACCGCAGTCGGA.....CAGCGTACCTTCGG 9936
578 rThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgProv 595
9935 CGCTTGGCGCACCGAGCTGGGTGGCGCTGACACATCGCACACTCGCAACGACTACT 9886
595 alPheProThrGlnGlyMetSerHisSerValAspLeuThrValGlyPhe 611
: : : : : |||||||..... : : : : : ||| : : : : :
9885 TCATGCCAAATCTTGTTATGATCAGCGCATTTGGCGCAGAGTGACACATA 9836
612 GlyAspLysThr.....HisGlnLysValValTyrGlnGlyAsnIleTyr 626
: : : : : : : : : : : |||||||..... : : : : : ||
9835 CCGCGCTCAACGATCAAGTACTACAAAATCAATACCAGATCTCGAAATA 9786
626 rArgProPheIleLysSerValLeuArgGlyTyrAlaLysLeuGlyT 643
|||||..... |||||||..... |||||||..... : : : : :
9785 CTGGCCAAATTATCCCGCTAGTACTCAACACGCGCTTGAAGTAGGT 9736
643 yrGlyAsnAsn..... 646
|||||.....
9735 ACGGACGACTACGSAAGAGTACACATCGGATCTGCTCGCGGACT 9686
647 .....LeuProPheTyrGluAsnPheTyrAlaGlyG 657
|||||..... |||||||..... |||||||..... |||||||.....
9685 GTGGCCACTGCTTCGGATTTGCTTTTCGAAAATTTCTACGCGCGGG 9636
657 yTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerG 674
: |||||||..... : : : : : |||||||..... |||||||.....
9635 CACCACTCGGTACAGGCTTCAGAGATAACACCTAGGACCGCGCTCGG 9586
674 lnaATyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluVal 690
: : : : : ||| : : : : : ||| : : : : :
9585 AA.....GTGACCGCAGCTGACAAATCAA.....GGCCAAACCA 9554
691 valGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuIleuProLe 707
|||||..... : : : : : |||||||..... : : : : :
9553 TTGGGTGGTTCCTTCAAGACTGTGGATCGACTGAAATGACTTCCCAA 9504
707 uProPheLysGlyAspTrpIleAspGlnValArgProValIlePheIleG 724
||| : : : : : ||| : : : : : ||| : : : : :
9503 GCTTTTCGACAGC.....CCATCGCGCAGCTATCTCAGCTTTCCTTG 9463
724 luGlyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIleAsp 740
: : : : : |||||||..... : : : : : ||| : : : : :
9462 ATTTCGGCAACGCTTCAAT..... 9443
741 LeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAl 757
: : : : : ||| : : : : : |||
```

```
9442 .....GGCGTAAATAATTTCAAGGC 9423
757 aAlaAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaG 774
: : : : : |||||||..... : : : : : ||| : : : : :
9422 A.....ACGAATTAGCTGCTTCAGCG 9400
774 lyValGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyr 790
|||||..... ||| : : : : : |||||||..... : : : : : |||
9399 CGGTAGCGCTGCTATGCGTGCCTCCCATTTGGTCCGATCTCGATCAGCTAT 9350
791 AlalysProLeuAsnLysGlnAsnAspGlnThrAspThrValGlnPh 807
||| ||| : : : : : |||||||..... : : : : : |||
9349 GCATTTCTATAAGAGAGAAACACGACGAAATCGAGCGACTGCAATT 9300
807 eGlnIleGlySerValPhe 813
: : : : : |||
9299 CACCTTCGCGCGCCCAATTC 9281
seq_name: gb_pat:AX155443
seq_documentation_block:
LOCUS AX155443 2379 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 6 from Patent WO0138350.
ACCESSION AX155443
VERSION AX155443.1 GI:14536771
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria. 1 to 2379)
REFERENCE
AUTHORS Giuliani,M.M., Pizsa,M., Rappuoli,R. and Holst,J.
TITLE 85kda neisserial antigen
JOURNAL Patent: WO 0138350-A 6 31-MAY-2001;
Chiron Spa (IT) ; Statens Institutt for Folkehelse (NO)
FEATURES
source
1..2379
/db_name="Neisseria meningitidis"
BASE COUNT 636 a 755 c 596 g 392 t
ORIGIN
alignment_scores:
Quality: 1123.50 Length: 837
Ratio: 2.073 Gaps: 19
Percent Similarity: 64.755 Percent Identity: 32.616
alignment_block:
US-09-701-711-2 x AX155443 ..
Align seg 1/1 to: AX155443 from: 1 to: 2379
10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
: : : : : : : : : : : |||||||..... : : : : :
4 AAACCTGAACAGATTGCTCCGCTCCGCTGATGATGTTGGGCATATCGCTTT 53
26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
|||||..... : : : : : ||| : : : : :
54 GGCAATTGCGGACTTCACCATCCAGACATCCGTCGGAAGGCTTGCGAC 103
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
||| : : : : : ||| : : : : : ||| : : : : :
104 GTACCGAGCGGACGACCGTATTCACCTACCTGCCCCCAAGTCGGCGAC 153
60 valValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaIath 76
: : : : : : : : : : : : : : : : :
154 ACCTACACGACACACACGCGAGTGCATCATCAAAAGCCTGTACGCCAC 203
76 rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIleT 93
||| ||| : : : : : ||| : : : : : |||
204 CGGTTTCTTTGACGAGTACGAGTTCGAAACTGCGGACGCGAGCTTCTGC 253
```


[illegible]


```

308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspSerGlyTyrTy 324
   :::::::::::::::::::::|||||
904 ACCGCGCTTTGGGTGAGATTAGAACCGGATGGCTCGCGAGGCTACGC 953
   :::::::::::::::::::::|||||
324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
   :::::::::::::::::::::|||||
954 ATACAGCGAAATCAGCGTACAGCGCTGCGCAACCGCGGAAACCAACCG 1003
   :::::::::::::::::::::|||||
341 alAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
   :::::::::::::::::::::|||||
1004 TCATTTCTGCTCGACATCGACCGGCGCGGAAATCTACGTCAACGAA 1053
   :::::::::::::::::::::|||||
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgG1 374
   :::::::::::::::::::::|||||
1054 ATCCACATCACCAGCAACAACAACCCGCGAGCGAGTGTGCGCGCGA 1103
   :::::::::::::::::::::|||||
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
   :::::::::::::::::::::|||||
1104 ATTGCGCAATGGATCCCGGCTTTACGACACCTCCAAAGCTGCAACGCT 1153
   :::::::::::::::::::::|||||
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
   :::::::::::::::::::::|||||
1154 CCAAGAGCGGCTCGAGCTTTGGCTACTTCGACACGCTACAGTTGAT 1203
   :::::::::::::::::::::|||||
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
   :::::::::::::::::::::|||||
1204 GCGTCCCGCTTGGGTGATGTCGCGCGGTATCGCAGGACACCTGTTCGGT 1253
   :::::::::::::::::::::|||||
424 lGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlnS 441
   :::::::::::::::::::::|||||
1254 GACGGAACGCTCCACCGGCTGCGTCTGACTTGAGCGCGGCTTCCAGG 1303
   :::::::::::::::::::::|||||
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
   :::::::::::::::::::::|||||
1304 ATACCGGCTTGGTTCATGTCGCGCGGTATCGCAGGACACCTGTTCGGT 1353
   :::::::::::::::::::::|||||
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
   :::::::::::::::::::::|||||
1354 ACGGCAAGTCGCGCGGCTGCGCGCTCGCGAAGCAAAACACCGCTCAA 1403
   :::::::::::::::::::::|||||
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
   :::::::::::::::::::::|||||
1404 CGGCTCGCTGCTGTATACGACCGGTACTTCAGCGCAGCGGGTCCAGC. 1452
   :::::::::::::::::::::|||||
491 lnSerLeuSerGlyTyr.....TyrArgLysThr.....LysTyr 502
   :::::::::::::::::::::|||||
1453 .....CTGGGCTAGATATTACGGAAGACCTTCGACCGCGGCAAA 1494
   :::::::::::::::::::::|||||
503 AspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerIle 519
   :::::::::::::::::::::|||||
1495 GCATCGACCAAGCTCAAAACAATATAAACCCACCACCGCGCGCGGCGT 1544
   :::::::::::::::::::::|||||
519 uSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536
   :::::::::::::::::::::|||||
1545 AAGGATGGGTATCCCGTTTACCGAATACGACCGCTCAATTTCCGGCTGG 1594
   :::::::::::::::::::::|||||
536 snAlaAspAsnThrLysLeuHisGly..... 544
   :::::::::::::::::::::|||||
1595 CGCGGGAACACCTGACCGTCAACACCTACAAACAAAGCACCAACGCTAT 1644
   :::::::::::::::::::::|||||
545 .....GlyArgPheMetGlyIleSerAsnVally 554
   :::::::::::::::::::::|||||
1645 GCCGACTTTATCAGGAATACGCGCAAAACCGACGCG..... 1680
   :::::::::::::::::::::|||||
554 sGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleP 571
   :::::::::::::::::::::|||||
1681 .....GCAGACGCG..... 1689
   :::::::::::::::::::::|||||
571 roAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuLeuGlyTyrAsn 587
   :::::::::::::::::::::|||||
1690 ..AGCTTCAAAGGC.....CTGCTGTACAAAGGACCGCTCGGCTGGGG 1731
   :::::::::::::::::::::|||||
588 TyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHisSe 604

```

```

1732 CGCAACAAGACCGACAGCGGTCTATGCGCGGCGGCTACCTGACCGG 1781
   :::::::::::::::::::::|||||
604 rValAspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysV 619
   :::::::::::::::::::::|||||
1782 CGTAATACGCGAAATCGCTCGCGCGGAGCAAACTGCAATACCTACCTCCG 1831
   :::::::::::::::::::::|||||
619 alValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeu 635
   :::::::::::::::::::::|||||
1832 CCACCCACACCAACCTGGTCTCTCCCTTAACCAAAACCTTCACGCTG 1881
   :::::::::::::::::::::|||||
636 Arg.....GlyTyrAlaLysLeuGlyTyrGly.....As 645
   :::::::::::::::::::::|||||
1882 ATGCTCGCGCGGAAGTCGCAATTCGCGGC...GGCTACGCGAGAACAA 1928
   :::::::::::::::::::::|||||
645 nasnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerVala 662
   :::::::::::::::::::::|||||
1929 AGAAATCCCTTCTTTGAAACCTTCTACGGCGCGGCTGGGTTCGGTGC 1978
   :::::::::::::::::::::|||||
662 rgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThr 678
   :::::::::::::::::::::|||||
1979 GCGGCTACGAAGCGGCACGCTCGCGCGGAA..... 2010
   :::::::::::::::::::::|||||
679 AlaArgArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyG1 693
   :::::::::::::::::::::|||||
2011 .....GTGTATGACGAATACGCGGCAAAATCAGCTACGCGCG 2048
   :::::::::::::::::::::|||||
693 yAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheL 710
   :::::::::::::::::::::|||||
2049 CACAAAGCAAGCAACGCTCGCGCGGAGCTCTCTCCGATGCCGCGGTG 2098
   :::::::::::::::::::::|||||
710 ysGlyAspTrpIleAspGlnValArgProValIlePheIleGluGlyGly 726
   :::::::::::::::::::::|||||
2099 CGAAGAC...GCACGACCGCTCGCGCTGAGCCTGTTGCGGACGCGAGC 2145
   :::::::::::::::::::::|||||
727 GlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGl 743
   :::::::::::::::::::::|||||
2146 AGCGTGTGG.....GACGCGAGAACCTATACCGCGCGCGCA 2180
   :::::::::::::::::::::|||||
743 nPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnA 760
   :::::::::::::::::::::|||||
2181 AACGGTAACAACAATCGGTTTACTCGGAACCGCGCAATAATCC.... 2226
   :::::::::::::::::::::|||||
760 rgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGly 776
   :::::::::::::::::::::|||||
2227 .....ACCTTTACCAACGAATTCGCTATTCCGCGCGCGCGCG 2265
   :::::::::::::::::::::|||||
777 AlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysPr 793
   :::::::::::::::::::::|||||
2266 GTTACCTGGCTCTCGCTTTGGTTCGATGAAATTCAGCTACGCTACCC 2315
   :::::::::::::::::::::|||||
793 oLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleG 810
   :::::::::::::::::::::|||||
2316 GCTGAAGAAAAACCGGAAGCAAAATCCAACGCTTCCAATTCAGCTCG 2365
   :::::::::::::::::::::|||||
810 lySerValPhe 813
   :::::::::::::::::::::|||||
2366 GCAGGAGTTC 2376
   :::::::::::::::::::::|||||
seq_name: gb_pat:AX155439
seq_documentation_block:
LOCUS AX155439 2394 bp DNA linear PAT 23-JUN-2001
DEFINITION Sequence 2 from Patent WO0138350.
ACCESSION AX155439
VERSION AX155439.1 GI:14536770
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 2394)
AUTHORS Giuliani,M.M., Pizza,M., Rappuoli,R. and Holst,J.

```

TITLE 85kDa neisserial antigen
 JOURNAL Patent: WO 0138350-A 2 31-MAY-2001;
 Chiron spa (IT); Statens Institutt for Folkehelsetse (NO)
 FEATURES Location/Qualifiers
 source 1..2394

/organism="Neisseria meningitidis"
 /db_xref="taxon:487"

BASE COUNT 657 a 743 c 583 g 411 t

ORIGIN

alignment_scores:

Quality: 1115.00 Length: 819
 Ratio: 2.042 Gaps: 14
 Percent Similarity: 66.667 Percent Identity: 32.479

alignment_block:

US-09-701-711-2 x AX155439 ..

Align seg 1/1 to: AX155439 from: 1 to: 2394

```

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
   :::::::::::::: ::::: |||:::|||||::: ::::
4  AAACGAAACAGATTGCTCCGCATGATGTTGGGCATATCGCCTTT 53
   |||:::|||||::: ::::: |||:::|||||::: ::::
26 aglnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
   |||:::|||||::: ::::: |||:::|||||::: ::::
54 GGCATCTGCCGACTTCACCATCCAGACATCCGCGTCGAAGCGTTGCAGC 103
   |||:::|||||::: ::::: |||:::|||||::: ::::
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
   |||:::|||||::: ::::: |||:::|||||::: ::::
104 GTACGAGCGGAGTAGTACCTATTCACTACCTGCCGCTCAAGTCGGCGAC 153
   |||:::|||||::: ::::: |||:::|||||::: ::::
60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76
   :::::::::::::: ::::: |||:::|||||::: ::::
154 ACCTACAGGACACACAGCGGATCCATCATCAAAAGCGCTGACGCCAC 203
   :::::::::::::: ::::: |||:::|||||::: ::::
76 rGlyAsnPheSerAspValGlnValTyrHisGlnGlyArgIleIleT 93
   |||:::|||||::: ::::: |||:::|||||::: ::::
204 CGGTTTCTTTCAGCAGTACGCGTGGGAACTGCGGACGGGACGTCCTGC 253
   |||:::|||||::: ::::: |||:::|||||::: ::::
93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109
   |||:::|||||::: ::::: |||:::|||||::: ::::
254 TGACGGTTATCGAACGCCGCCACCATCGGCTCGCTCAACATCACCGGCGCA 303
   |||:::|||||::: ::::: |||:::|||||::: ::::
110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126
   :::::::::::::: ::::: |||:::|||||::: ::::
304 AAAATGCTGCAAAACGACGCCATTAAAGAAAACCTCGAATCGTTCGGGT 353
   |||:::|||||::: ::::: |||:::|||||::: ::::
126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143
   |||:::|||||::: ::::: |||:::|||||::: ::::
354 GCGCAGTGCAGTAATCTTTAATCAGCGCACACTCAATCAGGCGATCGCGG 403
   |||:::|||||::: ::::: |||:::|||||::: ::::
143 LuLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr 159
   |||:::|||||::: ::::: |||:::|||||::: ::::
404 GCCTGAAGAAGAAATACCTCGGCGGCGGCAAACTCAATATCCAATCAGC 453
   |||:::|||||::: ::::: |||:::|||||::: ::::
160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176
   |||:::|||||::: ::::: |||:::|||||::: ::::
454 CCCAAAGTAACCAAACTCGCCCGCAACCGCTCGACATCGACATCAGCAT 503
   |||:::|||||::: ::::: |||:::|||||::: ::::
176 eAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG 193
   :::::::::::::: ::::: |||:::|||||::: ::::
504 TGACGAGGCGCAATCCGCGCAATACCCGACATCGAATTTGAAGCGCAACC 553
   |||:::|||||::: ::::: |||:::|||||::: ::::
193 InHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209
   |||:::|||||::: ::::: |||:::|||||::: ::::
554 AGTCTATTCCGACCGCAAACTGATCGGCAATATGTCCTGACCGCAAGGC 603
   |||:::|||||::: ::::: |||:::|||||::: ::::
210 LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLe 225
   |||:::|||||::: ::::: |||:::|||||::: ::::
604 GGCATTTGGACATGGCTGACACGAGCAACCAATTCACGAGCAGCAAAAT 653

```

```

225 uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV 242
   ::::: |||:::|||||::: ::::: |||:::|||||::: ::::
654 TGCCCAAGATATGGAAGAAATGTAACCGACTTCTACCAAAATTAACGCTACT 703
   |||:::|||||::: ::::: |||:::|||||::: ::::
242 alArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258
   |||:::|||||::: ::::: |||:::|||||::: ::::
704 TCGATTTCCTCGTATCTCATCCGACATCCAAACCAACGAAAGACAAACCC 753
   |||:::|||||::: ::::: |||:::|||||::: ::::
259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheGl 275
   ::::: |||:::|||||::: ::::: |||:::|||||::: ::::
754 AAGCAGACCATAAATCACCCTCCACGAAGCGGACGTTTCCGTTGGGG 803
   |||:::|||||::: ::::: |||:::|||||::: ::::
275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG 291
   :::::::::::::: ::::: |||:::|||||::: ::::
804 CAAAGTCTCCATCGAAGCGGACACCAACGAAGTCCCAAGCGGAACCTGG 853
   |||:::|||||::: ::::: |||:::|||||::: ::::
291 luAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeu 307
   |||:::|||||::: ::::: |||:::|||||::: ::::
854 AAAAATCTGCTGACCATGAAGCCCGCAATGGTACGAACGCCAGCATG 903
   |||:::|||||::: ::::: |||:::|||||::: ::::
308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyrTy 324
   :::::::::::::: ::::: |||:::|||||::: ::::
904 ACCGCCGTTTGGGTGATGATTCAGAACCGCATGGCTCGGAGGCTACGC 953
   |||:::|||||::: ::::: |||:::|||||::: ::::
324 rTyralaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
   |||:::|||||::: ::::: |||:::|||||::: ::::
954 ATACAGCGAAATCAGCGTACAGCGCTGCCGAACGCTGAAACCAAAACCG 1003
   |||:::|||||::: ::::: |||:::|||||::: ::::
341 alAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
   |||:::|||||::: ::::: |||:::|||||::: ::::
1004 TCGATTTCGCTGTCATCATCGAACCGCGGAAATCTACGTCAACGAA 1053
   |||:::|||||::: ::::: |||:::|||||::: ::::
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgGl 374
   :::::::::::::: ::::: |||:::|||||::: ::::
1054 ATACACATCACCGGCAACAAACAAACCGCGAGAGTCTGCCCGGTGA 1103
   |||:::|||||::: ::::: |||:::|||||::: ::::
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
   |||:::|||||::: ::::: |||:::|||||::: ::::
1104 ATTAGCCAAATGGAATCCGACCTTACGACACCTCCCAAGCTGCAACGTT 1153
   |||:::|||||::: ::::: |||:::|||||::: ::::
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
   |||:::|||||::: ::::: |||:::|||||::: ::::
1154 CCAAGAGCGCGTCAGCTTTGGGCTACTTCGACAATGTCAGTTTGTAT 1203
   |||:::|||||::: ::::: |||:::|||||::: ::::
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
   :::::::::::::: ::::: |||:::|||||::: ::::
1204 GGTGTCGCGCTTCCGCGCAGCGCGCAAGTCGATTTGAACATGAGTCT 1253
   |||:::|||||::: ::::: |||:::|||||::: ::::
424 lGluGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnS 441
   :::::::::::::: ::::: |||:::|||||::: ::::
1254 GACCGAAGCTTCCACCGGTTCCCTCGATTGAGCCGCGGTGGGTTCAAG 1303
   |||:::|||||::: ::::: |||:::|||||::: ::::
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
   :::::::::::::: ::::: |||:::|||||::: ::::
1304 ATACCGGGTGTGTCATGTCGCGAGCGGTTTCCCAAGACAACTGTTCCGT 1353
   |||:::|||||::: ::::: |||:::|||||::: ::::
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
   |||:::|||||::: ::::: |||:::|||||::: ::::
1354 ACGGCAAGTCCGCGCGCATCGCGCCCTCCAGGACCAAAACCCACGCTTAA 1403
   |||:::|||||::: ::::: |||:::|||||::: ::::
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
   :::::::::::::: ::::: |||:::|||||::: ::::
1404 CGGCTCGCTGCTGTTTACTACCGCTTACTTACGCGCAGACGCGGTCAGCC 1453
   |||:::|||||::: ::::: |||:::|||||::: ::::
491 InSerLeuSerGlyTyrTyrArg.....LysThrLysTyrAspAsnLys 505
   :::::::::::::: ::::: |||:::|||||::: ::::
1454 TGGGCTAGCATGTTTACGGAAGAGCTTCGACCCGCGCAAGCATCGACC 1503
   |||:::|||||::: ::::: |||:::|||||::: ::::
506 AsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrGl 522
   |||:::|||||::: ::::: |||:::|||||::: ::::
1504 AGCATCAACAATAATAAAACCAACCGGCGGCGGCGCATCCGATGAG 1553
   |||:::|||||::: ::::: |||:::|||||::: ::::
522 yTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539

```

[illegible]


```

2194 .....GGCGGCGGGTTCAAACATTTACCGCGCGCGCAATACCCA 2234
||||| ..... |||||
761 oLeuLeuThrGlnAspLysGlnIleuArgTyrSerAlaGlyValGlyAlaT 778
||||| ..... |||||
2235 TAAATCCACCTTTACCAAGAAATTCGCGTAATCCCGCGCGCGCGGTTA 2284
||||| ..... |||||
778 hTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
||||| ..... |||||
2285 CCGTGCCTCGCCTTTAGGCCGATGAATTCACCTACGCTACCGCTG 2334
||||| ..... |||||
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySer 811
||||| ..... |||||
2335 AAGAAAAACCGGAAGACGAATCAACGCTTCCCAATTCCAACTCGCGAC 2384
||||| ..... |||||
811 rValPhe 813
2385 GACGTTT 2391
seq_name: gb_ba:AE002375

```

```

seq_documentation_block:
LOCUS AE002375 13538 bp DNA linear BCT 25-MAY-2000
DEFINITION Neisseria meningitidis serogroup B strain MC58 section 17 of 206 of
the complete genome.
ACCESSION AE002375 AE002098
VERSION AE002375.1 GI:7225394
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis MC58.
Neisseria meningitidis MC58
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.

```

```

REFERENCE
1 (bases 1 to 13538)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Massignani, V.,
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
Rappuoli, R. and Venter, J.C.
Complete genome sequence of Neisseria meningitidis serogroup B
strain MC58
Science 287 (5459), 1809-1815 (2000)
20175755
10710307
2 (bases 1 to 13538)
Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C.,
Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F.,
Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D.,
Hickey, E.K., Haft, D.H., Salzberg, S.L., White, O., Fleischmann, R.D.,
Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E.,
Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H.,
Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Massignani, V.,
Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R.,
Rappuoli, R. and Venter, J.C.
Direct Submission
Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

```

```

FEATURES
source
1. .13538
/organism="Neisseria meningitidis MC58"
/strain="MC58"
/db_xref="taxon:122586"
/note="serogroup: B"
complement(121. .1377)
/gene="NMB0176"
complement(121. .1377)
/gene="NMB0176"
/note="similar to GB:L02948 SP:P29011 PID:145703 GB:U00096
PID:1651591 percent identity: 76.98; identified by
sequence similarity; putative"

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS

```

gene

```

gene
CDS
/note="similar to GB:L02948 SP:P29011 PID:145703 GB:U00096
PID:1651591 percent identity: 76.98; identified by
sequence similarity; putative"

```

```

/codon_start=1
/transl_table=11
/product="D-amino acid dehydrogenase, small subunit"
/protein_id="AAF40633.1"
/db_xref="GI:7225395"
/translation="MKVLVLGAGVAGVSSAWYLAEAGHEVTVVIVDRAEGVAMETSFNANA
GQLSYGYTTTAAAPGIPTRKALKWLFKSHPLLPFRPDGSLYQIEMWQMLQNCQTAARYO
TKNRWVRISSEYREMFREFAQTGMNFGRRKGTQIFQRTKEVAARQDQAVLARY
GVYRRLKPEECAEPALARTAKTAGGLHLPADATDCRFLTENLYLCOEKGVRF
HENQNLSDHNGLRKTYETETGREADAVVCAALGCFSEFTVLAQDLNLPYVKGX
SLTLPVNSDGPVSTLVDESXYATREDNRIRVCGMAELSGYEKLPEKRETLAL
VYNDLFPPEGDLQSQAFWSGLRPMTDPSTPLIGRTFRDNLFLNTGHTGLGWTMSLGS
KLTDIVSGKTEIRSDLSLSRYQA"
complement(1456. .2847)
/gene="NMB0177"
complement(1456. .2847)
/gene="NMB0177"
/note="similar to GB:AE000511 PID:2314080 percent
identity: 69.11; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="sodium/alanine symporter, putative"
/protein_id="AAF40634.1"
/db_xref="GI:7225396"
/translation="MQVFDNPKAFETVSGWVWGPLMLMLLVGTGILLTVLLKGLQF
TMLGVALKOARFMPKPKKSGEGHGDISHFAALMTALSATIGTNTAGVATAVVTGSP
GAVFWNMTAIFGMATKYCEGYLVAKYRVNNSKSGSGPMYIEKGLGNWKMVA
FALGTFASGIVSSVQSAQAVQTSFGIEPAYTGITLVLTAVVGLGKIGIARA
ASFIVPMAVYVVLGLSLIAINSDALMPAVKLIFDSAFSAQAVAGAIGTVIRYGA
GVFVNEAGMSAPIAAAAKTDHPVRQALVSTGFTLDTIVVCSITGIVLMGLLGA
GGFVYKPEVSGAALTVTTFQKMLPGIGGWITVIGLFFAYSTILGCWYCEKCAVVF
GEKFAGLYRVGVYSSVMLCTVLSLDLVLWLASDTFNGLMALPNLIALLMLAKVIVNETR
DFKQKITNGELPH"
complement(3204. .3980)
/gene="NMB0178"
complement(3204. .3980)
/gene="NMB0178"
/note="similar to PID:1718489 percent identity: 100.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl-(acyl-carrier-protein)--UDP-N-
acetylglucosamine O-acetyltransferase"
/protein_id="AAF40635.1"
/db_xref="GI:7225397"
/translation="MTLIHPTAVIDPKAELDSGVKGVATVIGPNVQICANTEIGHPA
VINGHTSIGENNRITQFASLGEIPQDKYRDEPTKLIIGNGNTIREFTTNLGTVTGI
GETRIGDDNIMAYCHLADCVIGNHTIFANNASLAGHVTIGDYVVLGYTLVVFQCR
IGDYAMTAPAGVHKDVPYFMAAGYRAEPAGLNSGEMRRNGFTABQISAVKDVYKTL
YHRGIPFEAKADILRRATQAEALAVFRDFAQSARGIIR"
complement(4067. .4516)
/gene="NMB0179"
complement(4067. .4516)
/gene="NMB0179"
/note="similar to PID:1718488 percent identity: 100.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="(3R)-hydroxymyristoyl-(acyl carrier protein)
dehydratase"
/protein_id="AAF40636.1"
/db_xref="GI:7225398"
/translation="MDVQLPIEAKDTOKLIPHRYPFLQDRITAEPMKTLTAIKNVT
INEPQDGHFPDLVPVPGVLIIEAQAACGTAILISEGGRKENEFFFGAGIDEARFR
QVIPGQDLFEVLELTSTRIGKIGNAVAKVDGQVAVEAIIKARVV"
complement(4550. .5596)
/gene="NMB0180"
complement(4550. .5596)
/gene="NMB0180"
/note="similar to PID:1718487 percent identity: 100.00;
identified by sequence similarity; putative"
/codon_start=1

```

```
/transl_table=11
/product="UDP-3-O-(3-hydroxymyristoyl)-glucosamine
N-acetyltransferase"
/protein_id="AAF40637.1"
/db_xref="GI:7225399"
/transl_table=11
/transl_table="MIPATVLSQITARLGEWRGEDI SVTVRPLADAQAEHISFLA
NPKYAEVHSSAGAVIYSAKADGFEGRNLIVADDPILYFAKVARLFSFVVRKAGGI
HPTAVFPGATVPTSCBEGANVYIGANTVLGECRILANAVQHDCLVDEVLHPNA
VYVGTGLGRVEIHSAGVADGFGFLAFADSWFKIPOTGAVTIGDDVDVIGNTNID
RGMSDITVNGNKTIDNOVQIGHNCKIGSHTVIAAKTIGISGVTIGSYCIIGGVGVTV
GHIFADKTTIGGTSYTHSITESKHLAGIFPMSTHKEWARNVYIHRLESMNRKLIK
TLEQLSDAQDQSK"
/complement(5629. .6129)
/gene="NMB0181"
/complement(5629. .6129)
/gene="NMB0181"
/notes="similar to GB:M34854 SP:P31519 PID:155479 percent
identity: 46.88; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein OmpH, putative"
/protein_id="AAF40638.1"
/db_xref="GI:7225400"
/transl_table="MTRLRAFAALIGLCTAGAHADTFQKGFINTERIYLESKOA
RKQKTDSEFSARQDELQLOREGLDLRLAEGLRNKAKKAEEKRWGLVAAFRK
KQAFEDRYNLRNEEPASLQONANRVIVKIQEGYDVILQNVYVNTQYDVTDSVI
KEMNAR"
/complement(6195. .8588)
/gene="NMB0182"
/complement(6195. .8588)
/gene="NMB0182"
/notes="similar to GP:2460281 percent identity: 99.87;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein Omp85"
/protein_id="AAF40639.1"
/db_xref="GI:7225401"
/transl_table="MKLQJASALMIGISPLADFTIODIRVEGLQRTPESTVPVNY
LPVKGYDYNTHGSAIILSYXAGFEDDYRVETADGOLLITVIERPTIGSLNKGAK
MLQNDALKNLESFGLAQSYFNQATLNQAVAGLKEYLGRGKLNQIITPKVTKLARN
RVDIDITDEGSAKIDTIEFEGNQVYSDRKLMLQMSLTGEGITWLTNRNQNEQRF
ADQMEKTYFQNNGYDFRLDITQTNEDKTQTIKTIVHEGGRWKGVSIEGDT
NEVPKAELEKLLTMKPKQYERQOMTAVLGEIQRNMGSAQYSEISVQPLPNAETKT
VDFVLHTEPKRTIVNEIHIHTGNKTRDVRRELRQMESAPYDTSKLORSKERVELL
GYFNQVQDAPVPLAGTDPKVDLNMSLTERSTGSLDLSAGVQDTGLVMSAGVSQDNLF
GTGSAALRASRKTTLNGLSLFTDPTFTADGVSGLDVGKAFDPKASTISKQYKT
TTAGALIRMSVPVTEYDRVNFGLVVAEHLTVNTYNKPKHYADFIRKYGTGDTGDSFK
GWLYKTVGWRNKTDLSALPRTGYLTGVNAEIALPGSKLQYTSATHNOTWFFPLSKT
FTMLGEGVGTAGYGRTKETPEPFENFYGGGLGSRVYEGSTGLGPKVYDYGKISYG
GNKKANYSALLRPPMCAKDARTVRLSLFADAGSVWDGKTYDDNSSSATGRVQNIYG
AGNTHKSTFTNELRYSGAGVATWLSPLGPMKFSVAYPLKKKPEDEIQRFPQLGTFP"
/complement(8645. .9985)
/gene="NMB0183"
/complement(8645. .9985)
/gene="NMB0183"
/notes="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF40640.1"
/db_xref="GI:7225402"
/transl_table="MHTLLAFIFAILLIVSLHEFGHVIVARLCGVKVVRFVSGFGKPF
FTRKRGTECLAPIPLGCVGVKVDTRFEGVSEADLPYAFDKQHPAKRIALVAGPLT
NLALVLLYGLSFVGTVELRPYVETPIDAARAGFQSGDKIQSVNGVPVADWGA
QTEVLNLEAGVAGVQVTSAGQVETIDAAGTPEAGKAKNQGYIGLMLKITTYA
GVGKSPGSAKGLPKDRLTAAADKPIASQEWANLTRQSPGKKTILNTERAGQHTT
ADIRPDVQESDHTLIGRGLRQPDRAWDQAIIRSVYRPSVRFAGMWEKTVSHSWT
TLKFFGKLISGNASVSHISGLPIADIAGOSAEIGLOSYLEFLALVSIIGVLNLLPV
PVLDDGHLVPTTAEIRGKPLGERVQNIIGRLRFLGLAMLMMAVAFNDVTRLLG"
/complement(10020. .11204)
```

```
/gene="NMB0184"
/complement(10020. .11204)
/gene="NMB0184"

alignment_scores:
  Quality: 1115.00      Length: 819
  Ratio: 2.042         Gaps: 14
  Percent Similarity: 66.667  Percent Identity: 32.479

alignment_block:
US-09-701-711-2 x AE002375/rev ..

Align seg 1/1 to reverse of: AE002375 from: 1 to: 13538

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8585 AAACCTGAACACAGATGCTTCCGCACTGATGATGTTGGGCATATCGCCTTT 8536
26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8535 GGCACCTTGGCGACTTCACCATCCAAGACATCCGCGTCGAAGGCTTGACGC 8486
43 rGValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8485 GTACCGAGCGGAGTACCGATTCACTCCGCGTCAAAAGTCGCGGAC 8436
60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8435 ACCTACACGACACACAGCGCAGTCCCATCATCAAAAGCCTGTACGCCAC 8386
76 rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIleThr 93
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8385 CGGTTTCTTTGACGAGTACGCGTCGAAACTCGGACGGCGGACGCTCCTGC 8336
93 yrGluValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8335 TGACGCTTATCAACGCCGCCACATCGCTCGCTCAACATCACCGCGGCA 8286
110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8285 AAAATGCTGCAAAACGACGCCATTAAAGAAAACCTCGAATCGTTCGGCT 8236
126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8235 GCGCAGTCGCAATACTTTTAATCAGCGCACACTCAATCAGGACGTCGCGG 8186
143 LuLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr 159
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8185 GCCTGAAAGAGAAATACCTCGGCGCGGCAAACTCAATATCCAAATACAG 8136
160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8135 CCCAAGTAACCAAACTCGCCCGCAACCGCTCGACATCGACATCAGAT 8086
176 eAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG 193
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8085 TCACGAGGCAAAATCCGCAAAATCACCACATCGAAATTTGAAGGCAACC 8036
193 InHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
8035 AAGTCTATTCCGCGCAAACTGATCGGCAAAATGTCCCTGACCGAAGGC 7986
210 LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLe 225
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
7985 GCATTTCGACATGCTGACACAGAACCACTCAACAGGACAGAAAT 7936
225 uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV 242
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
7935 TGCCCAAGATATGAAAAAGTAGTACCGACTTCTTACCAAAATAACGCGTACT 7886
242 aIArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258
   ::::::::::: ::::::::::: ::::::::::: ::::::::::: :::::::::::
7885 TCGATTTCGATCTCTCTACCGACATCCAAACCAACGACAGCAAAACC 7836
```

```

259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheG1 275
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7835 AAGCAGACCATCAAAATCACCGTCAGAGGCGGACGTTTTCCGTTGGG 7786
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG 291
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7785 CAAAGTCTCCATCGAAGCGGACACCAAGAGTCCCCCAAGCCGAACGTG 7736
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
291 luAlaLeuLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeu 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7735 AAAACGCTGTGACCATGAAGCCGGGAAATGGTACGAACGCCACGATG 7686
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
308 GluGlnThrThrAsnAniLeSerThrLysPheGlyAspAspGlyTyrTy 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7685 ACCCGCTTTGGGTGAGATTCAAGCCGATGGCTCGCGGAGCTACGC 7636
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7635 ATACAGCGAAATCAGCTACAGCGCTGCCGAACGCTGAAACCAAAACCG 7586
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
341 alaSpValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7585 TCGATTTCCGTCGACATCGAACCGCGCGGAAATCTACGTCAACGAA 7536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgG1 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7535 ATACACATCACCAGCAACAACAAACCCGCGACGAAGTCTGCCCGCTGA 7486
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7485 ATTACGCCAAATGAATCCGACCTTACGACACACCTCCAAAGCTGCAAGTT 7436
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
391 erArgAlaArgLeuMetArgThrClyPhePheLysHisValThrValAsp 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7435 CCAAAGAGCGCGTGGAGCTTTGGGTCTTTCGACAATGTCACGTTTGAT 7386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7385 GCTGTCCCGCTTGGCGCAGCGCGACAAAGTCGATTTGAACATGAGTCT 7336
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
424 lGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlnS 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7335 GACCGAACGTTCCACCGGTTCCCTGGATTTGAGCGCGGTTGGGTTCAAG 7286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7285 ATACCGGTTGGTCATGTCGCGAGCGCTTTCCCAAGACACCTGTTGCGT 7236
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7235 ACGGCAAGTCCGCGCACCTGCGCGCTCCAGGAGCAAAACACGCTTAA 7186
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
474 lTyrSerGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7185 CGCTGCTGCTGTTACTGACCGTACTTTCACGGCAGCGGGTTCAGCC 7136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 lnSerLeuSerGlyTyrTyrArg.....LysThrLysTyrAspAsnLys 505
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7135 TGGGTACGATGTTACGGAAGAGCCTTCAGCCCGCGCAAGCATCGACC 7086
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
506 AsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrG1 522
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7085 AGCATCAACAATATAAAACCAACCCAGCGGACGCGGCGCATCCGCTAG 7036
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
522 yTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7035 CGTGCCTGTACCGAATACGACCGCTGATTTTCGTTTGGTGGCAGAAC 6986
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6985 ACCTGACCGGTCAAC.....ACCTACAACAAGCGCCCAAA 6951
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: gb_pat:AX043922

seq_documentation_block:

LOCUS

AX043922

linear

PAT 24-NOV-2000

DEFINITION Sequence 1 from Patent WO0066791.


```
458 ThrGlyLysHisValAsnAlaSerPheArgSerGluThrArgGluVa 474
|||||...:|
179453 ACGGCAAGTCGGCCACATGCGCGCTCCAGAGCAAAACACCGCTTAA 179404
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
|||||...:|
179403 CGGCTCGCTGCTGTTACTGACCGCTACTACGGCAGAGCGGGTTCAGCC 179354
491 lnsrLeuSerGlyTyrTyrArg.....LysThrLysTyrAspAsnLys 505
...:|
179353 TGGGCTACCATGTTTACGAAAGCCTTCGACCGCGCAAGCATCGACC 179304
506 AsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrG 522
...:|
179303 AGCATCAACATATATAAACACCACCGGAGCGGCAGGATCCGCATGAG 179254
522 yTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
:|||||...:|
179253 CGTGCCTGTTACCGAATACGACCGCTGAATTCGGTTTGGTGCGAGAAC 179204
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
...:|
179203 ACCTGACCGCTCAAC.....ACCTACAAACAAAGCGCCCAA 179169
556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAs 572
|||||...:|
179168 CACTATGCGGACTTTATCAAGAATAACGGCAAAACCGACGGC....ACAGA 179122
572 pPheLysHisAspTyrThrTyrAsnAlaIleLeuGlyTyrAsnTyrS 589
|...:|
179121 CGCGACTTCAAAAGCTGGCTGTACAAAGGTACCGTCCGGTGGGGCGCA 179072
589 eSerLeuAspArgProValPheProThrGlnGlyMetSerHisServa 605
...:|
179071 ACAAACCCACACGCGGTATGCGCCACCGCGGCTGACCTACGCGGCGG 179022
606 AspLeuThrValGlyPhe...ThrHisGlyLysValVa 620
...:|
179021 AACGCGGAATCGCCCTGCTGGCAGCAAACTGCAATACTACTCCGCCAC 178972
620 lTyrGlnGlyAsnIleTyrArgProPheIleLysLysServaLeuArg. 636
...:|
178971 CCACACCAACACCTGTTCTCCCTGAGCAAAACCTTCACGCTGTATGC 178922
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
178921 TCGGCGCGGAAGTCGGCATTCGGGCG...GGCTACGGCAGAACCAAGAA 178875
647 LeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgG 663
...:|
178874 ATCCCTCTCTTTGAAAACCTCTACGCGCGCGGCTGCGGTGCGTGGCGG 178825
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAla 680
|||||...:|
178824 ATACGAAACGGCAGCCTCGGTCGGA..... 178797
680 rgArgGlyGlnGlnThrLeuGlyGluValVal.....GlyGlyAsn 694
|||||...:|
178796 .....GTCATGACGAATACGGCGAATAATACAGTACGGCGGCAAC 178755
695 AlaLeuAlaIlePheGlySerGluLeuIleLeuProLeuProPheLysG 711
|||||...:|
178754 AAAAAAGCCACGCTCTCCGCGAGCTGCTTCCCGATCCCGCGCGCGAA 178705
711 yAspTrpIleAspGlnValArgProValIlePheIleGlyGlyGlnV 728
|||||...:|
178704 AGAC...GCGCGCACCGCTCGCGCTGAGCGCTGTTGCGGACGAGGCGAG 178658
728 aPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
|||||...:|
178657 TGTGGACGCGCAAAACCTACGACGACAAACAGAGATTCGCGGACCC..... 178614
```

```
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
...:|
178613 .....GGCGGACGGTTCAAACATTTACGGCGCGCGCAATACCCA 178573
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
|||||...:|
178572 TAAATCCAGCTTTACCAACGAATTCGCTATTCCGCGCGGCGCGGTTA 178523
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
|||||...:|
178522 CTRGGCTCTCGCCTTTAGGCCCGATGAATTCAGCTACGCTACCCGCTG 178473
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
...:|
178472 AAGAAAAACCGGAGACGAAATCCACGCTTCCAATTCCACTCGGCAC 178423
811 rValPhe 813
...:|
178422 GACGTTC 178416
seq_name: gb_pat:AX155447
seq_documentation_block:
LOCUS AX155447 2394 bp DNA linear PAT 22-JUN-2001
DEFINITION Sequence 10 from Patent WO0138350.
ACCESSION AX155447
VERSION AX155447.1 GI:14536772
KEYWORDS
SOURCE Neisseria gonorrhoeae.
ORGANISM Neisseria gonorrhoeae
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE 1 (bases 1 to 2394)
AUTHORS Giuliani M.M., Pizsa M., Rappuoli R. and Holst J.
TITLE 85kda neisserial antigen
JOURNAL Patent: WO 0138350-A 10 31-MAY-2001;
Chiron Spa (IT); Statens Institut for Folkehelse (NO)
FEATURES
source 1. 2394
/organism="Neisseria gonorrhoeae"
/db_xref="taxon:485"
BASE COUNT 655 a 758 c 586 g 395 t
ORIGIN
alignment_scores:
Quality: 1111.00 Length: 819
Ratio: 2.035 Gaps: 14
Percent Similarity: 66.667 Percent Identity: 32.357
alignment_block:
US-09-701-711-2 x AX155447 ..
Align seg 1/1 to: AX155447 from: 1 to: 2394
10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
...:|
4 AAACCTGAACAGATTCGTTCCGACATGATGGTCTGGGCATATCGCCTTT 53
26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
|||||...:|
54 GGCACATTGCGGACTTCACCATCCAGACATCCGCGTCGAAGGCTTGCGAG 103
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
|||||...:|
104 GTACCGAGCGGATACCGTATTCAACTACCTGCGCTCAAGTCGCGGAC 153
60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76
...:|
154 ACTACACGACACACGCGAGTGCATCATCAAAGCCTGATACGCCAC 203
76 rGlyAsnPheSerAspValGlnValTyrHisGlnGlyArgIleIleT 93
|||||...:|
```

[illegible]

```

680 rgargGlyInsgInThrThrLeuGlyValVal.....GLYGLYAsn 694
2011 .....GTGATGACGAATACGGGAAAAATCAGCTACGGCGGCAC 2052
695 AAlaLeuAlaThrPheGlySerGluLeuLeuProLeuProPheYsgl 711
2053 AAAAAAGCCACAGTCGCCGCCGAGCTGCTTCCGATGCCCGCGCGAA 2102
711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGln 728
2103 AAcG...GGCGCACCGCTCCGCTGAGCTGTGGCCGACGCGAGCGC 2149
728 aAlpheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
2150 TGTGGACGGCAAAACCTACGACGACACAGCAGTTCCGCGACC..... 2193
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAsnArgpr 761
2194 .....GGGGCGAGGTTCAAAACATTTACGGCGCGCAATACCCA 2234
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 778
2235 TAAATTCACCTTTACCAACGAATGGCTATTCGCCGGCGCGCTTA 2284
778 hTrpThrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
2285 CCTGGCTCTCGCTTACGCCGATGAATTCAGCTACGCTACCGCGCTG 2334
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGly 811
2335 AAGAAAAACCGGAAGACGAATCCAAACGCTTCATTCCACTCGCGAC 2384
811 rValPhe 813
2385 GAGCTTC 2391

seq_name: gb_pat:AX202497
seq_documentation_block:
LOCUS AX202497 2394 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 10 from Patent WO0152885.
ACCESSION AX202497
VERSION AX202497.1 GI:15392217
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 2394)
AUTHORS
Plaza,M.C., Rappuoli,R.C. and Giuliani,M.C.
TITLE
Outer membrane vesicle (omv) vaccine comprising N. meningitidis
serogroup b outer membrane proteins
Patent: WO 0152885-A 10 26-JUL-2001;
JOURNAL
Chiron Spa (IT)
FEATURES
source
1..2394
Location/Qualifiers
BASE COUNT 655 a 758 c 586 g 395 t
ORIGIN
1..2394
/organism="Neisseria meningitidis"
/db_xref="taxon:487"

alignment_scores:
Quality: 1111.00 Length: 819
Ratio: 2.035 Gaps: 14
Percent Similarity: 66.667 Percent Identity: 32.357

alignment_block:
US-09-701-711-2 x AX202497 ..
Align seg 1/1 to: AX202497 from: 1 to: 2394
10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26

```

```

.....: .....: |||.....:
4 AAACGTAAACAGATGCTCTCCGACATGATGTCTTGGCATATACGCTTT 53
26 aglnAlaAlaAspPheMetAlaAsnAspIleThrIleThyLeuGlnA 43
54 GGCACCTTGTCCGACTTCACCATCCAGACATCCCGTCGAAGCGTTGAGC 103
43 rGValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGln 59
104 GTACCGAGCCGAGTACCGATATTCACACTACTGCGCGTCAAGTGGGAGC 153
60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAla 76
154 ACCTACAGACACACACAGCGCATGCCATCATCAAAACCTGTACGCGAC 203
76 rGlyAsnPheSerAspValGlnValTyrHisGlnGlyArgIleIle 93
204 CGGTTCTTGTGACGAGTACGCTGCAAACTGCGACGCGCACCTCTCG 253
93 yrglnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109
254 TGACCGTTATGACAGCGCCCATCATCGCTGCTCAACATCACCGCGCA 303
110 ArgLeuIleProLysGlnGlyLeuGlnGlyLeuLysAsnAlaGly 126
304 AAATGCTGCAAAACGACGCCATTAAGAAAACCTGCATCTGTGGGCT 353
126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThr 143
354 GCGCGAGTCGCATCTTATCATCAGGCGACATCATCAAGCAGTGGCGG 403
143 lLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr 159
404 GCCTGAAGAAAGATACCTCGGGCGGCGCAACTATATCAATATCAG 453
160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThr 176
454 CCCAAAGTAAACCAACTCGCCCGCCGACCGGCTGACATCCATCAACAT 503
176 eAlaGluGlyLysProAlaArgValAlaAspIleAsnIleIleGlyAsn 193
504 TGACGAGGGCGAAATCCGCAAAATCAACCGCATCTGAATTTGAAGCGAAC 553
193 lHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209
554 AACTCTATTCCGACCGCAAACTGATGCGGAGATGTCTGACCGAAGGC 603
210 LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLys 225
604 GGCATTTGGACATGGCTGACACGACGACCAATTCACAGCAGCAAAAT 653
225 uValThrSerLeuGlnAsnLeuArgAlaLysTyrLeuAsnAlaGlyPhe 242
654 TGCCCAAGACATGAAAGTAACCGACTCTACCAAGAACACAGCGTACT 703
242 aAlaPheGluIleLysAspAlaLysLeuAsnIleAsnGlnLysAspAsn 258
704 TCGATTTCCGATCTCGATACGACATCCAAACGAAAGCAAGCAAAAC 753
259 ArgIlePheValGluIleSerLeuHisGluGlyGluIleTyrArgPhe 275
754 AAGCAGACCATCAAAATCACCGTCCACGAAAGCGGAGCTTCCGTTGGG 803
275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeu 291
804 CAAGTCTCATGGAAGCGACCAACGAAAGTCCCAAGCGCAAGCTGG 853
291 lAlaLeuLeuLysPheLysAlaGluGluGlnLysSerGlnAlaMetLeu 307
854 AAAAAGCTGCTGACCATGAAGCCCGGCAAAATGTACGAAAGCCAGAGAT 903
308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspLysTyr 324
.....: .....: |||.....:

```

904 ACCGCCGTTTGGGTGAGATTTCAGAACCGCATGGCGGCGAGCTACGC 953
 324 TTYTAlaGlnIleArgProValThrArgIleAsnAspGlnSerArgThrV 341
 954 ATACAGCGAAATCAGCTACAGCCGCTGCCAAGCCGCAACCAACCAACCG 1003
 341 aAlaSpValGlnUtyrThrIleAspProValHisProValTyrValArg 357
 1004 TCGATTTCGCTCGCATCGAACCGGCGCGGAATCTACGTCACGA 1053
 358 IleAsnPheThrGlyAsnPheYsrThrGlnAspGluValIleuArgG 374
 1054 ATCCACATCCCGGCAACACCAAAACCGGACGAGTGTGCGCGCGCA 1103
 374 uMetArgGlnLeuGlnGluAlaIleuAlaSerAsnGlnIleGlnLeu 391
 1104 ATTGCGCCAAATGGAATCCGGCGCTTACGACACCTCCAAAGTCGACG 1153
 391 eArgAlaArgLeuMetArgThrGlyPhePheYsrHisValThrValAs 407
 1154 CCAAAGACGCGCTGAGCTTTGGCTTCTGACACGTCACGTTGTGAT 1203
 408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheVal 424
 1204 GCGTCGCGCTGGCGGACACCGCAAGTCGATTTGACATGACGCT 1253
 424 IGluglnInProSerGlySerSerThrIleAlaIleArgIleYsrGln 441
 1254 GACCGCACTTCCACCGGCTGCTGCTGCTGAGCGCGGCTGGTACAGC 1303
 441 eArgGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
 1304 ATACGCGCTGTGTCATGTCGAGCGGTTTCCAGACAACTGTTCGT 1353
 458 ThrGlyHisValAsnAlaSerPheSerArgSerGluThrArgGlnVal 474
 1354 ACGGCAAGTCGCGCGCGCTGCGGCTCAGCAACCAACCAACCTCA 1403
 474 ITyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSer 491
 1404 CGGCTCGCTGTACCGACCGCTACTTCAACGCGACGCGGTCAGGC 1453
 491 InSerLeuSerGlyTyrTyrArg....LysThrLysTyrAspAsnLys 505
 1454 TGGGCTACAGTGTTCGGAAGGCTTCGACCGCGCAACCAATCGAC 1503
 506 AsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrG 522
 1504 AGCATCAACATATATAAACCAACCAACGCGACGCGATCCGATGAG 1553
 522 YThrProIleAspGlnAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 539
 1554 CGTGCTGTACCGAATACGACGCGGTGAATTCGTTGGTGGCAGAAC 1603
 539 SerThrLysLeuHisGlyLysArgPheMetGlyIleSerAsnValLysGln 555
 1604 ACCTGACCGCTCAAC.....ACCTACACAAACACCGCCCAAA 1638
 556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIlePro 572
 1639 CACTATGCGCGATTTATCAAGAAATACGCAAAACCGAGGCG...AC 1685
 572 PheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrPAsnTyr 589
 1666 CGGCAAGCTTCAAGGCTGCTCTACAAAGTACCGTGGGCGGCGCA 1735
 589 eArgLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
 1736 ACAAAACGACAGCGCTTATGCGGCGGCGCTACTACGAGGCGGCG 1785
 606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVal 620
 1786 AACGCGGAATCGCTGCGCGCGCAAACTGCAATCTACTCGCGCAC 1835

620 ITyrGlnGlyAsnIleTyrArgProPheIleLysSerValLeuArg. 636
 1836 CCACAAACCAACCTGCTTCTTCCCTTAAGCAAAACCTTACAGCTGARGC 1885
 637GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
 1886 TCGCGCGGCAACCTGCGCATTTGGCGGC...GGCTACGGCAAGAACCAAA 1932
 647 LeuProPheTyrGluAsnPheTyrAlaGlyIleTyrGlySerValArg 663
 1933 ATCCCTTCTTGAATACTTCTACGCGCGGCTGCGGCTGCGTGGCGG 1982
 663 YThrAspGlnSerLeuGlyProArgSerGlnAlaTyrLeuThrAla 680
 1983 ATACAAAGCGGACGCTGCTGCGCAA..... 2010
 680 rArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
 2011GTGTATACGAATATGCGCAAAATATAGCTACGCGCGCAC 2052
 695 AlaLeuAlaThrPheGlySerGlnLeuIleLeuProLeuProPheLysG 711
 2053 AAAAAAGCAAGCTCTCGCGGAGCTGCTTCCGATGCGCGGCGGCA 2102
 711 YAspThrIleAspGlnValArgProValIlePheIleGlnGlyGlnV 728
 2103 AGAC...GGCGCACGCTCGCGCTGAGCTGTTCGCGACGAGGACAG 2149
 728 aPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
 2150 TGTGGAGCGCAAACTACGACGACACAGCATTCGCGCAC..... 2193
 745 LysAspProGlnAlaThrAlaGlnGlnAsnAlaLysAlaAsnArgPr 761
 2194GGCGGACGAGGTCAAAACATTTACGCGCGCGGCAATACCA 2234
 761 OleuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 778
 2235 TAAATCCACCTTACCAAGAAATTCGCTATTCGCGCGGCGGCTTA 2284
 778 hTriPyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
 2285 CCTGGCTCGCTTATGAGCCGATGAATTCAGCTACCGCTACCGCTG 2334
 795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGly 811
 2335 AAGAAAAACCGAAGACGAATCAACCTTCCAAATTCCACTCGGCGAC 2384
 811 rValPhe 813
 2385 GACGTTG 2391
 seq_name: gb_ba:NM12491
 seq_documentation_block: 340806 bp DNA linear BCT 04-DEC-2000
 LOCUS NM12491
 DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome;
 segment 1/7.
 ACCESSION AL162752 AL157959
 VERSION AL162752.2 GI:7378778
 KEYWORDS
 SOURCE
 ORGANISM
 Neisseria meningitidis Z2491.
 Neisseria meningitidis Z2491
 Bacteria; proteobacteria; beta subdivision; Neisseriaceae;
 Neisseria.
 1 (bases 1 to 340806)
 REFERENCE
 1 Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
 Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
 Davies,R.M., Davis,P., Devlin,K., Felwell,T., Hamlin,N.,
 Holtrodt,S., Jagers,K., Leather,S., Moule,S., Mungall,K.,
 Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
 Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.

TITLE Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491
JOURNAL Nature 404 (6777), 502-506 (2000)
MEDLINE 20222556
REFERENCE 2 (bases 1 to 340806)
AUTHORS Parkhill, J.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2000) Submitted on behalf of the Neisseria sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
COMMENT Notes:
 Details of *N. meningitidis* sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).
FEATURES

source	1..340806	/organism="Neisseria meningitidis Z2491"
		/strain="Z2491"
		/db_xref="taxon:122587"
gene		/note="serogroup: A"
		complement(38..358)
CDS		/gene="NMA0001"
		complement(38..358)
		/gene="NMA0001"
		/codon_start=1
		/product="hypothetical protein NMA0001"
		/protein_id="CAB83321.1"
		/db_xref="GI:7378779"
		/db_xref="SPTREMBL:O9JX93"
		/translation="MSNTQIRSADEFTTSRAWGALDIANMNGTIVRLHWTDDQPKWHV NDGEVFAVMGDEVMDHYRENGEEHIVRLKSGDIFVAGICTEHAHPGRGTRILVIEK EGSV"
gene		complement(414..2438)
CDS		/gene="nuoL"
		complement(414..2438)
		/gene="nuoL"
		/EC_number="1.6.5.3"
		/note="NMA0002, nuoL, NADH dehydrogenase I chain L, len: 674 aa; similar to many e.g. NUOL_RHOCA NADH dehydrogenase I chain L (EC 1.6.5.3) (700 aa). Contains Pfam match to entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone (complex I), various chains, and to entry PF00662 oxidored_q1.N, NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus"
		/codon_start=1
		/transl_table=11
		/product="NADH dehydrogenase I chain L"
		/protein_id="CAB83322.1"
		/db_xref="GI:7378780"
		/db_xref="SWISS-PROT:O9JX92"
		/translation="MNDMTLYLIALVPLAGSLIAGLFGNKGIGRAGHTVTILGVAVS AVLSAYVLGFLNGSRKEDENYTWLTMGLOFSGFLVDMTAMMMVVTVGSVMV HRTTQMDKRGVGRFYSISLFTSLMLIMNSNFIOLFGESEAVGLVSLYLGIF YFRSATLANKALINRDFGLGILVLAIFGSLRDQDFAYLPNVQNAITQ LFPGWMSITVCLLTFVGMGSAQFPLHVLWPDMSMEGPTISALIIHAATVMTAGL FVYRMSPIYEMSSTALSIWVIGATALFMGLVIQNDIKRWVAYSTISOLGYMTV ALGASAYSVAMPHVMTAFKALLFLAAGSAIIGMHDDQDMRHGNLKKYMPITWLT LIGNLSLIGTFPFGYSKDSIIIEAAKYSTLPSCGFAYFAVLASVFVTAFAFRQFM VFGEKRWSLPEHSDGHEHGLGKNDNPESPLVTLPLILLAVSPVITIGIYAI EPLVIGDFKDVIFVNADAPHTMHIMKEFHGALAMVSHSLSPVLYLALAGVLSAWL LVKPLHLPKTAQAFRPVYVLPENRYLDALYFNVFAKGRALGTFFFWKVGDTAIIID NGIVNGSARLVGAQAQVRKVQTFYTYAAAMVFGVLVLLGTFWGLFR"
misc_feature	1035..1044	/note="Core DNA uptake sequence: gccgtctgaa"
		/label=DUS
misc_feature		complement(1116..2030)
		/gene="nuoL"
		/note="Pfam match to entry PF00361 oxidored_q1, NADH-Ubiquinone/plastoquinone (complex I), various chains, score 351.80, E-value 7.3e-102"
misc_feature		complement(2061..2243)

gene		/gene="nuoL"
		/note="Pfam match to entry PF00662 oxidored_q1.N, NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus, score 77.20, E-value 3.5e-19"
		complement(2487..2828)
		/gene="NMA0003"
		complement(2487..2828)
		/gene="NMA0003"
		/note="NMA0003, unknown, len: 113 aa"
		/codon_start=1
		/transl_table=11
		/product="hypothetical protein NMA0003"
		/protein_id="CAB83323.1"
		/db_xref="GI:7378781"
		/db_xref="SPTREMBL:O9JX91"
		/translation="MMDKNOLEQEFHFKAMLIYOEALNLPQYKATRFLOIVNEFGGK EAADKLLSTGCKTQTGTETLILSGGVHALKSMYELVLQKPCWDLFTBQLAVARK RLERVGFVSPK"
		complement(2833..2837)
		complement(2952..3527)
		/gene="NMA0004"
		complement(2952..3527)
		/gene="NMA0004"
		/note="NMA0004, unknown, len: 191 aa; similar to hypothetical proteins e.g. Y977_HAEIN HI0977 (191 aa), fasta scores: E(): 0; 80.5% identity in 190 aa overlap, and to proteins involved in cell cycle e.g. TR-054679 (EMBL:AF036487) Lactococcus lactis Plasmid pNZ4000 putative mobilization protein (200 aa), fasta scores: E(): 0, 55.1% identity in 167 aa overlap, and FIC_ECOLI cell filamentation protein FIC (200 aa), fasta scores: E(): 0.093, 29.4% identity in 126 aa overlap"
		/codon_start=1
		/transl_table=11
		/product="hypothetical protein NMA0004"
		/protein_id="CAB83324.1"
		/db_xref="GI:7378782"
		/db_xref="SPTREMBL:O9J0R9"
		/translation="MPSENPIGTKMSIDFQSLNARLPESGDIDRIVCTAGLQO IHRYLGGLYDFAGQIREDNISKGFRFANMYLKEALVKEQNPETFEIILAKYVE MNALPPELVGNGRSRIWLDVLNKKVYVQNVSKTLYLQAWERSPVNDLEIRFL LKDNLTDDVDNREIIFKIQSYIEGYEGK"
		complement(3516..3525)
		/gene="NMA0004"
		/note="Core DNA uptake sequence: gccgtctgaa"
		/label=DUS
		complement(3561..3866)
		/gene="nuoK"
		complement(3561..3866)
		/gene="nuoK"
		/EC_number="1.6.5.3"
		/note="NMA0005, nuoK, NADH dehydrogenase I chain K, len: 101 aa; similar to many e.g. NUOK_RHOCA NADH dehydrogenase I chain K (EC 1.6.5.3) (102 aa), fasta scores: E(): 1.1e-24; 65.0% identity in 100 aa overlap. Contains Pfam match to entry PF00420 oxidored_q2, NADH-ubiquinone/plastoquinone oxidoreductase chain 4L"
		/codon_start=1
		/transl_table=11
		/product="NADH dehydrogenase I chain K"
		/protein_id="CAB83325.1"
		/db_xref="GI:7378783"
		/db_xref="SPTREMBL:O9J0U9"
		/translation="MITLTHVLVIGALLFGISAMGIFMRNKNVLVLMSTELMLAVN FNIAPSOHLGDTAGQIFVFVLTVAANESAIGLAINVLYVRNQTINVADLDELK"
		complement(3564..3863)
		/gene="nuoK"
		/note="Pfam match to entry PF00420 oxidored_q2, NADH-ubiquinone/plastoquinone oxidoreductase chain 4L, score 141.50, E-value 1.5e-38"
		complement(3863..4534)
		/gene="nuoJ"
		complement(3863..4534)

```

/EC_number="1.6.5.3"
/Note="NMA0006, nuoJ, NADH dehydrogenase I chain J, len:
223 aa; similar to many e.g. NUOJ, RHOCA NADH dehydrogenase
I chain J (EC 1.6.5.3) (202 aa), fatsa scores: E():
9, 7e-21; 41, 3% identity in 201 aa overlap. Contains Pfam
match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase I chain J"
/protein_id="CAB83326.1"
/db_xref="GI:7378784"
/db_xref="SPTREMBL:Q9UX90"
/translation="MTFGILFIYFAVILYGAIKTVAKNPVHAALHLVTFVCSAM
LWMLMQAEFLGVLTVVVFVGVAMVLFVVMNLNIDIEEMRAGFWRHAPVAGVVGTL
AVALLILVNPKTDLAAFLMKMDIPADYNNIRDLGSRITYDYLLPELAAVLLLGWV
RAIALVHRKTVNPKRMDPADQVKVPRADQGRMLVKMEAVKQTESAESEVSDDLKPK
EEGKA"
RBS
complement(3874..3879)
/gene="nuoJ"
/misc_feature
complement(4034..4525)
/gene="nuoJ"
/Note="Pfam match to entry PF00499 oxidored_q3,
NADH-ubiquinone/plastoquinone oxidoreductase chain 6,
score 77.90, E-value 2e-19"
/gene="NMA0007"
complement(4570..5277)
/gene="NMA0007"
complement(4570..5277)
/Note="NMA0007, unknown, len: 235 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0007"
/protein_id="CAB83327.1"
/db_xref="GI:7378785"
/db_xref="SPTREMBL:Q9UX89"

```

alignment_scores:

```

Quality: 1111.00 Length: 819
Ratio: 2.035 Gaps: 14
Percent Similarity: 66.667 Percent Identity: 32.357

```

alignment_block:

```

US-09-701-711-2 x NMA1Z2491 ..

```

```

Align seg 1/1 to: NMA1Z2491 from: 1 to: 340806

```

```

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
83985 AAACGTGAACAGATGTCTCCGACATGATGGCTTGGGCATATCGCCTTT 84034
26 aGlnAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
84035 GGCACCTTGGCGACTTCACCATCCAGACATCCGCGTCGAAGGCTTGCG 84084
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59
84085 GTACCGAGCGAGTACCGGTATTCACATACCTGCGCCGTCAAAGTCGGCG 84134
60 ValValSerGluAsnGlnLeuAlaAspGlyValIlysAlaLeuValAla 76
84135 ACCTACACGACACACAGCGAGTGCCATCAAAAGCCGTGACGCCAC 84184
76 rGlyAsnPheSerAspValGlnValTyThrHisGlnGluGlyArgIleIle 93
84185 CGGTTTCTTGACGACGTACGCGTGAAGCTGCGGACGGCGAGCTCTCG 84234
93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109
84235 TGACCGGTATTCGAACGCCCCACCATCGCTCGCTCAACATCACCGGGCA 84284
110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126

```

```

84285 AAATGTCGCAACAGCGCATTAAGAAAACCTCGAATCGTTTCGGGCT 84334
126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143
84335 GCGCAGTCGCAATACTTAAATCAGGCGACACTCAATCAGCAGCTCGCG 84384
143 luLeuThrAsnGlnTyIleSerGlnGlyTyTyAsnThrGluIleThr 159
84385 GCCTGAAGAAGAAATACCTCGGGCGGCAAACTCAATATCAATCAAC 84434
160 ValIysGlnThrMetLeuAspGlyAsnArgValIlysLeuAspMetThr 176
84435 CCCAAATGATACCAAACTCGCCGCAACCGCTCGACATCGACATCAGAT 84484
176 eAlaValGlyGlyProAlaArgValAlaAspIleAsnIleIleGlyAsn 193
84485 TGACGAGGGCAATCCGCCAAATCACCAGACATCGAATTTCAAGGCAAC 84534
193 lnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209
84535 AAGTCTATTCCGACCGCAAACTGATGGCGAGATGTGCTGACCGAAGGC 84584
210 LysIle...AsnProLeuSerLysAlaAspArgTyThrGlnGluLysLe 225
84585 GGCATTTGGACATGGCTGACGACGAAGCAACCAATTCACAGCAGCAAA 84634
225 uValThrSerLeuGluAsnLeuArgAlaLysTyThrLeuAsnAlaGlyPhe 242
84635 TGCCCAAGACATGGAAGAAAGTAACCGACTTCTACCAAGACACACGCT 84684
242 alArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258
84685 TCGATTTCCGCATCTCTGATACCGACATCCAAACCAAGCAAGCAAAAC 84734
259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyArgPheG 275
84735 AAGCAGACCATCAAAATCACCGTCCACGAGGCGGCGCTTTCCGTTGGG 84784
275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyThrGlnAlaGluLeuG 291
84785 CAAGTCTCCATCGAAGCGCACACCAAGTCCCAAGCCCAAGCAACTGG 84834
291 luAlaLeuLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeu 307
84835 AAAAATCTGTGACCATGAGCCCGGCAATGGTACGAACGCCACGACATG 84884
308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyTy 324
84885 ACCGCCGTTTTGGGTGAGATTACAGACCGCATGGGCTCGGCGAGCTAGC 84934
324 rTyAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
84935 ATACAGCGAAATCAGCGCTACAGCGCTGCCAAAGCCGCAACCAACCG 84984
341 alAspValGluTyTyIleAspProValHisProValTyValArgArg 357
84985 TCGATTCGCTCGCATGCAACCGCGCGGAAATCTACGTCAACGAA 85034
358 lIeAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgG 374
85035 ATCCACATACCGGCAACACAAACCCGCGGACGAGTCTGTCGCCCGCGA 85084
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeu 391
85085 ATTGCGCAAAATGGAATCCGCGCTTACGACACCTCCAAAGCTCAACGCT 85134
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
85135 CCAAGAGCGCGCGAGCTTTGGGCTTCTCGACAAACGACGATACGTTGAT 85184
408 ThrArgProValProAsnSerProAspGlnValAspValAspValVal 424

```

```
85185 GCGTCCCGCTGTCGCGCACACCCGACAAAGTCGATTGTAACATGAGCCT 85234
424 lGluGluInProSerGlySerThrIleAlaAlaGlyTyrSerGlnS 441
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85235 GACCGAAGCTTCCACGGCTCGCTGACATTGAGCGGGCTGGGTACAGG 85284
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
441 erGlyGlyValThrPhePheAspValSerGlnAsnAsnPheMetGly 457
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85285 ATACCGCGCTGGTCATCTCCGAGGCGTTTCCCAAGACACCTGTCGGT 85334
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85335 ACGGGAAGTCGCGCGCTTCGCGGCTCAGCAAGCAAAACACCGCTCAA 85384
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
474 lTyrSerLeuGlyMetThrAsnProThrPheThrValAsnGlyValSerG 491
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85385 CGCTCGCTGCTGTTTACCGACCGTACTTCACGGCAGCGGGTCAGCC 85434
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
491 lnSerLeuSerGlyTyrTyrArg.....LysThrLysTyrAspAsnLys 505
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85435 TGGGCTACGATGTTTACGGAAGAGCCTTCGACCCGCGCAAGCATCGACC 85484
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
506 AsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrGln 522
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85485 AGCATCAACAATATAAAACACCAACCGCGGCGGCGGCGCATGAG 85534
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
522 yTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85535 CGTGCCTGTACCATACGACCGGTGATTTCCGTTTGGTGGCAGAAC 85584
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85585 ACCTGACCGGTCAC.....ACCTACAACAAGCGCCCAA 85619
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnGlyIleProAs 572
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85620 CACTATGCGGACTTTATCAAGAATAACGGCAAAACCCGCGC...ACAGA 85666
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
572 pPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrS 589
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85667 CGCGACCTCAAGGCTGGCTGTACAAAGGTACCGTGGGTGGGGCGCA 85716
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
589 erSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85717 ACAAAACCGACAGCGGTTATGCGCGCGGCTACCTACCTGCGGGCGTG 85766
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85767 ACGCGGAATCGCCCTGCGCGGCGCAACAACTGCAATACTACTCCGCCAC 85816
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
620 lTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeuArg. 636
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85817 CCACAACCAACCTGTTCTTCCCTTAAGCAAAACCTTCACCGCTGATGC 85866
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85867 TCGCGCGCGAAGTCGGCATTCGCGGC...GGCTACGCGCAGAACCAAGAA 85913
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
647 LeuProPheTyrGluAsnPheThrAlaGlyGlyTyrGlySerValargGl 663
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85914 ATCCCTCTCTTTGAAACCTTACGGCGCGCGCTGGGTTCGGTGGCGGG 85963
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaA 680
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85964 ATACGAAGCGGCGACCTCGGTCCGAAA..... 85991
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
680 rArgGlyGlnGlnThrThrLeuLeuGlyGluValVal.....GlyGlyAsn 694
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
85992 .....GTGTATGACGAATACGGCGCAAAATACAGTACGCGCGCAAC 86033
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
695 AlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLysGl 711
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
86034 AAAAAAGCCAACCTCTCCGCGGAGCTGCTCTCCCGATGCCGCGCGGAA 86083
: ||||: :|||: :|||: :|||: :|||: :|||: :|||: :|||: :|||:
```



```
1504 AGCATCAAAACAATATATAAAACCACGAGCGAGCGCATCGCATGAG 1553
522 yTyProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
1554 CGTGCCTGTACCAATACAGCGCGTGAATTCGGTTTGGTGGCAGAAC 1603
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
1604 ACCTGACCGCTCAAC.....ACCTACAAACAAGCGCCCAAA 1638
556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAs 572
1639 CACTATGCCGACTTATCAAGAATAACGGCAAAACCGACGCG...ACAGA 1685
572 pPheLysHisAspTyThrThrTyAsnAlaIleLeuGlyTyrAsnTyrs 589
1686 CGCGACGCTCAAGGCTGGCTGTACAAAGGTACCGTCCGCTGGGGCGCA 1735
589 erSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
1736 ACAAAACCGACAGCGGTTATGCGCGACGCGGCTACCTGACGCGCGTG 1785
606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620
1786 AAGCGCGAAATCGCCTGCCGCGCAACAACCTGCAATACTACTCCGCCAC 1835
620 lTyRGlnGlyAsnIleTyArgProPheIleLysLysSerValLeuArg. 636
1836 CCACACCAACCTGGTCTTCCCTTAGCAAAACCTTCACGCTGATGC 1885
637 .....GlyTyAlaLysLeuGlyTyrgly.....AsnAsn 646
1886 TCGCGCGGAAGTCGCGCATGCGGCG...GGCTACGCGCAACCAAGAA 1932
647 LeuProPheTyrcLusAsnPheTyAlaGlyGlyTyrglySerValargGl 663
1933 ATCCCTCTCTTGAANAACCTTACGCGCGCGGCTGGGTTCGCTGCGCGG 1982
663 yTyAspGlnSerSerLeuGlyProArgSerGlnAlaTyrlLeuThrAlaA 680
1983 ATACGAAAGCGGACCGCTCGGTCCGAAA..... 2010
680 rgArgGlyGlnGlnThrThrLeuGlygluValVal.....GlyGlyAsn 694
2011 .....GTGTATGCAATAACGCGCAAAATAACAGTACGCGCGCAAC 2052
695 AlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLysGl 711
2053 AAAAAAGCAACGCTCTCCGCGAGCTGCTCTCCGATGCTCGCGCGCAA 2102
711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlnV 728
2103 AGAC...GGCGGACCGCTCGCGCTGAGCCTGTTCGCGGCGGAGGAGCG 2149
728 alphaAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
2150 TGTGGGACGCGCAAACTACGACGCAACAGCAGTTCGCGGACG..... 2193
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
2194 .....GGCGGACGGGTTCAAAACATTTACGCGCGCGCAATACCCA 2234
761 oLeuLeuThrGlnAspLysGlnLeuArgTySerAlaGlyValGlyAlaT 778
2235 TAAATCCACTTTACCAACGAATTCGCTATTCGCGCGCGCGCGGTGA 2284
778 hrTrpTyThrProIleGlyProLeuSerIleSerTyAlaLysProLeu 794
2285 CTTGGCTCTCGCCTTTAGGCGCGCAATTCAGTACGCTACCGCTACCGCTG 2334
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
2335 AAGAAAAACCGGAAGACGAAATCCAAACGCTTCCAATTCACACTCGGCAC 2384
```

811 rValphe 813

:::||||

2385 GACGTT 2391

seq_name: gb_ba:NGU81959

seq_documentation_block:

LOCUS NGU81959 2379 bp DNA linear BCT 24-JAN-2002

DEFINITION Neisseria gonorrhoeae outer membrane protein (omp85) gene, complete

cds.

ACCESSION U81959

VERSION U81959.1 GI:1766041

KEYWORDS

SOURCE

ORGANISM

Neisseria gonorrhoeae.

Neisseria gonorrhoeae

Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;

Neisseria.

REFERENCE 1 (bases 1 to 2379)

AUTHORS Manning,D.S., Reschke,D.K. and Judd,R.C.

TITLE Omp85 proteins of Neisseria gonorrhoeae and Neisseria meningitidis

are similar to Haemophilus influenzae D-15-Ag and Pasteurella

multocida Oms87

Microb. Pathog. 25 (1), 11-21 (1998)

MEDLINE 98379445

PUBMED 9705245

REFERENCE 2 (bases 1 to 2379)

AUTHORS Reschke,D.K., Manning,D.S. and Judd,R.C.

TITLE Direct Submission

Submitted (11-DEC-1996) Division of Biological Sciences, University

of Montana, Health Sciences 104, Missoula, MT 59812-1002, USA

FEATURES

Location/Qualifiers

1..2379

/organism="Neisseria gonorrhoeae"

/strain="FAL9"

/db_xref="taxon:485"

1..2379

/gene="Omp85"

1..2379

/gene="omp85"

/note="Omp85"

/codon_start=1

/transl_table=11

/product="outer membrane protein"

/protein_id="AAC17600.1"

/db_xref="GI:1766042"

/translation="MKLQIASALMMLGISPLAFADFTIQDIRVGLQTERSTVPNY

LPVKVGDYNDTHGSAIIKSATYATGFFDDVRVETADGLLLLTVIVCPFTIGSLNITGAK

MLQNDAIKKNFESGLAQSOYFNQATLNAQVAGLKEEYLGRLKLIQTTPKVTKLARN

RVDIDITIDEKSAKITDIEFGNOVYSDRKLQMSLTGGIWTWLTNRSDRFDRKF

AQDMKVTDFYQNGYFDRLDITDQNEKTRQITKTVHEGGRFWGKYSIEGDT

NEVPKAELEKLLTMKPGKWEBOOMTAVLGEIONRMGSAGYAYSEISVQPLNACTKT

VDFVLHIEPGRKIYNEIHTGNKTRDEVYRRELQMSAPYDTSKLQSKERVELL

GYFDNVQFDVPLAGTDPKVDLNMSLTERSTGSLDSAGVQDQTLGVMSAGVSQNLFF

GTGSAALRASRSKTLNLSLFTDPTFYTAGVSLGYDIYKAFDPKASTSVQYKT

TTAGGVRMGIPVTEYDRVNFGLAAEHLTVNTYKAPRYADFIKQYKTDGADSEFK

GLLYKGTVGWRNKTDSALPWTPTGYLTGVNBAIPLPGSKLOYYSATHNOTWFFPLSKT

FTLMLGGEVGIAGGRTKEIPFFENYGGGLGVSRYGSESTGLPKVVDYEKESYG

GNKANVSAEELFPMPKADARTVRLSFADAGSVWDGRTTYTAENGNNKSVYSENAH

KSTFTNELRYSAGGAVTWLSPLGPKMFIYAPLKKKPEDEIQRFQFLGTTT"

BASE COUNT 636 a 752 c 592 g 399 t

ORIGIN

alignment_scores:

Quality: 1104.50 Length: 822

Ratio: 2.038 Gaps: 17

Percent Similarity: 65.937 Percent Identity: 32.603

alignment_block:

US-09-701-711-2 x NGU81959 ..

Align seg 1/1 to: NGU81959 from: 1 to: 2379


```
1777 ACCGGCGTAAATGCGAAATCGCCTCGCCGCGCAAACTGCAATACTA 1826
      ::::::::::: ::::::::::: |||:::|
617 nLysValTyrGlnGlyAsnIleTyrArgProPheIleLysLysServ 634
      ::::::::::: ::::::::::: |||:::|
1827 CTCGCCACCCACCAACCACTGGTCTTCCCTTAAGCAAAACCTTCA 1876
      ::::::::::: ::::::::::: |||:::|
634 alLeuArg.....GlyTyrAlaLysLeuGlyTyrGly... 644
      |||
1877 CGCTGATGCTCGCGGCGAAGTCGCGATGCGGC...GGCTACGGCAGA 1923
      |||
645 ...AsnAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySe 660
      ::::::::::: |||:::|
1924 ACCAAAGAAATCCCTCTCTTTGAACACTTCTACGCGCGCGCTGGGTC 1973
      |||
660 rValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrL 677
      |||
1974 GGTGCGCGGTACGAAGGCGCAGCTCGGCCGAAA..... 2010
      |||
677 euThraLarArgGlyGlnGlnThrThrLeuGlyGluValVal..... 691
      |||
2011 .....GTGTATGACGAATACGCGCAAAATCAGCTAC 2043
      |||
692 GlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuPr 708
      |||
2044 GCGCGCAACAAAAAGCAACGCTCGCGCGAGCTGCTTCCCGATGCC 2093
      |||
708 oPheLysGlyAspTrpIleAspGlnValArgProValIlePheIleGluG 725
      |||
2094 CGGTGCGAAGAC...GCACGACCGTCCGCGCTGAGCTGTTGCCGACG 2140
      |||
725 lyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeu 741
      |||
2141 CAGCGACGCTGGG.....GACGCGCAGAACCTATACGCC 2175
      |||
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaI 758
      |||
2176 GCCGAAACGGAACAAACATCGGTTTACTCGGAAACGCGCATAAATC 2225
      |||
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
      |||
2226 C.....ACTTTACCACGAATTCGCTATTTCGCCGCGC 2260
      |||
775 alGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
      |||
2261 GCGCGTTACTGCTCTCGCTTTCGCGCGATGAAATTCATCATCGCC 2310
      |||
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
      |||
2311 TACCCGCTGAAGAAAAACCGGAAGACGAATCCAAGCTTCCAATCCA 2360
      |||
808 nIleGlySerValPhe 813
      |||
2361 GCTCGGCACGAGTTC 2376
      |||
```

seq_name: gb_ba:AF329831

seq_documentation_block:

LOCUS AF329831 2382 bp DNA linear BCT 03-JUL-2001
DEFINITION Haemophilus ducreyi integral outer membrane protein (DI5) gene,

complete cds.

ACCESSION AF329831

VERSION AF329831.1 GI:14586744

KEYWORDS

SOURCE

ORGANISM

Haemophilus ducreyi.

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

Haemophilus.

1 (bases 1 to 2382)

Thomas, K.L., LeDuc, I., Olsen, B., Thomas, C.E., Cameron, D.W. and

Elkins, C.

REFERENCE

AUTHORS

TITLE

Cloning, Overexpression, Purification, and Immunobiology of an
85-kilodalton Outer Membrane Protein from Haemophilus ducreyi

JOURNAL Infect. Immun. 69 (7), 4438-4446 (2001)
MEDLINE 21295095
PUBMED 11401984
REFERENCE 2 (bases 1 to 2382)
AUTHORS Thomas, K.L., LeDuc, I., Olsen, B., Thomas, C.E., Cameron, D.W. and
Elkins, C.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2000) Medicine, Univ. of N. Carolina, Chapel
Hill, NC 27599, USA
FEATURES
source
1..2382
/organism="Haemophilus ducreyi"
/db_xref="taxon:730"
1..2382
/gene="DI5"
1..2382
/gene="DI5"
/codon_start=1
/transl_table=11
/product="integral outer membrane protein"
/protein_id="AAK70345.1"
/db_xref="GI:14586745"
/translation="MKLLLSLLVANGVIAFPVVKDIRIDGVQTETGNATIASLPV
KVGQTADGDVTNIVKQFSONRFENVSAAREGQTLIVKVAERPLINNLTIKGNNAIP
KNALEQNKANLIVAGEVYDKAKLEAFKQALVDHYTMGRYQADIQTITPPNNGSIN
VELNITEGEIAYVAKINFGNNAFSYDELIKELEIKPNAPWNIFFSSKFOQOYKND
IELTRDFYMDHGYAKFTLKDQVKNENKTEVDLTVKINEGSOYNISEMRIIGDTOKL
DNELNQLLTHFKAGQLFRKTELSIIIEQIKQILGDRGYSKAVDLYPKPNEEDHTVQI
NFIVDAGRRIVVKIRFEGNDVTADSLRREMQQEGAMLSAVLSAKSRLEBRTGY
ETVEMSPVTVKNTDQVDIYIKERTGTSINFGVYSGSGSLYNAGITQDNFLMG
SSGLNGSRNTDSTNVNLSYTEPFTKDGVS LGNFIYEDYDINSARKASAAKYKRKYG
ASGTLGFPVDENNYSVLGYTHDKLRNVREYTRKYVNSMKFFPINPQNSHYDRIOS
ADFDLSFGWNNLNRGVPTAGSSANISGKLTPLGSDNKYVOYGTNFCYIPLNSPH
KWVIATKGGLAYTNSFGGKEVPYQILYSAGGMSLRGAGGSIGPKAIYRYREGFKAP
SQDVIGGNAMVNASLEIIPAPISDIQHNVRTSVFVDAATVWNTKWKSKADYPML
PDFGDYKRVRASAGIALQWQSPIGLSFSYAKPIKKIYAGDEIEQFOFTVGSTP"

BASE COUNT 821 a 414 c 447 g 700 t

ORIGIN

alignment_scores:

Quality: 1058.50 Length: 807

Ratio: 1.967 Gaps: 15

Percent Similarity: 66.667 Percent Identity: 31.599

alignment_block:

US-09-701-711-2 x AF329831 ..

Align seg 1/1 to: AF329831 from: 1 to: 2382

28 AlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgVa 44
||||| ::::::::::: ||| ||| ::::::::::: |||
49 GCGCACCATTTGTAGTAAAGATATTCGCATTGATGTTGTTCAACAGA 98
||| ::::::::::: ::::::::::: ||| ||| ::::::::::: |||
44 lThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGlnVal 61
||| ::::::::::: ::::::::::: ||| ||| ::::::::::: |||
99 AACTGGCAATGCAATTATTGTCATCTCTACCGGTTAAAGTTGGTCAAACTG 148
||| ::::::::::: ::::::::::: ||| ||| ::::::::::: |||
61 alSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGly 77
::: ::::::::::: ::::::::::: ||| ||| ::::::::::: |||
149 CAACAGATGGAGATGTTACTAATATTGTTAAACAACTTTTGTCAAAAC 198
||| ::::::::::: ::::::::::: ||| ||| ::::::::::: |||
78 AsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIleTyrGl 94
::: ||| ::::::::::: ||| ::::::::::: |||
199 CGTTTGTAGATGTAGTGTGCTGCGAGGCAACACATTAGTCATTAA 248
||| ::::::::::: ||| ||| ::::::::::: |||
94 nValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgL 111
||| ::::::::::: ||| ||| ::::::::::: |||
249 AGTTGCTGAACGCTCTTTGATTAACAATTTAACTATTAAAGAAACAATG 298
||| ::::::::::: ||| ||| ::::::::::: |||
111 eutleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAla 127
||| ::::::::::: ||| ||| ::::::::::: |||
299 CAATCCCTCAAAATGCAATAGAACAAAACCTTAAAGGCAAAATCTAATGTG 348
||| ::::::::::: ||| ||| ::::::::::: |||


```

gene
  GDVTFGSASLONKHQAEKLTENNNTALTNTAANYGGCWDIVQATKSLAAQOVKRGKLA
  VDEINAOVLQALVTKQPOVLLIRTSGEQRISNFIWQIAYAEIYFSDVLPDFDNE
  KFEENAIYQORHRRFGAEE"
  5479..6348
  /gene="cdsA"
  /note="PM1990"
  5479..6348
  /gene="cdsA"
  /codon_start=1
  /transl_table=11
  /product="CdsA"
  /protein_id="AAK04074.1"
  /db_xref="GI:12722430"
  /translation="MLKRVLSAIIAIVFIALFFSPFYFALAVGVVILAVWEWT
  QPAHFKFSFWRIGLTISSAAFLEWLYGESDYLNAGRFFENTLSLLFSSVLMWFVAL
  GLVTSYPKGATYWKGNLLOFAPAFLLVPFFAGVLRRLAHVYVTDVPHGLILLVYF
  ILVWSADGAVFACROFGKKNLAPKVPSPGKSWEGVIGGLVYTAGLILAFIFHTQESLL
  SGISIPAFIALSVATVAVSILGDLTESMFRQSGIKOSSQLIPGHGILRIDSLTAA
  VPFAYIFYFVIS"
  6370..7698
  /gene="PM1991"
  6370..7698
  /gene="PM1991"
  /codon_start=1
  /transl_table=11
  /product="unknown"
  /protein_id="AAK04075.1"
  /db_xref="GI:12722431"
  /translation="MSFLWFSFASFIIVISLVAVHEXGHEFWAARKGCIQVRRSIGRG
  KVLMSRDKOGTEFVISAIPLGYYVKMLDGRNEVVPPELSSRAFDQKSVLQRAFVIAH
  GPTANFELAILAFTIYTGIPVKKPIADISNSIAAQAIQEPNTOIMAVDGTGVSD
  WETINMLATKMGNDLHLTSPFGSIEQHKVNLTKDMRFDPEKESAMSSGLQIPVR
  FSDVILSVKVSADKAGKADRIYAGEQLISMQQFVQFVQEGKPPFNKVRERDQO
  FSVFVLPENLKKRWTVGTAAPISDIYRTKYLKYLEALQKVEKTIQLSWLT1
  KVIKFTGLDLKLNKGPISIAKGAGISSEIGLIYVYLGFMALISVNLGIMNLEPLPV
  LDGGHLVFLAAEAVRGKPLSERIQNLSYRIGAAIILMALMGFALENDFRL"
  7728..10103
  /gene="PM1992"
  7728..10103
  /gene="PM1992"
  /codon_start=1
  /transl_table=11
  /product="unknown"
  /protein_id="AAK04076.1"
  /db_xref="GI:12722432"
  /translation="MKLLIASLLFGSTTAAAPFVVKDIRVDGVQAGTEGSLATLP
  VRVQQRATNDIANVVRKFLSLGQYDDVKASREGNTLVVTMPKPVISNVVLDGNKSI
  PDEAIKONLDNFKVGDVNLRAKLEEFKRGIVHYNSVGRYNAKVEAIVNTLPNNSA
  EIKIQUEDDVALFKETIEGNOAFSSKLEDOMELQTDWKKLFGNKKFTQTFQNDL
  ETLSYLLDRGYAQFQILTDI KLSDDKKEARVLIKVEGDLTYVSARILGVDGMS
  AELAPILDAIQNLGLFRANVLEVEQRIKSKLGERGYATAQVNVHPFDPQDKTISLD
  FIVEAGKSVTRQIRFEGNTSSADSTLRQMRQOEGAWLSSELVELGKLRDRTPFE
  SVETKEAIPGSDQVDVYVYKERNTGSIINFGIGYGTESGLSYQASIKODNFLGMSG
  ISLGGTRNDYGTINLGNBPYTKDQVSLGNGVFEEDYDSSKSTSAAGRTSYGNG
  LTGLFPNNNSYLLGVGYTYNKLNAPEYNDLYROSMKYNDSTFKSHDRLSFG
  WYNSLNRGVPFKGRANIGCRVTIPGSDNKYKLNAAQGFYPLDRHGWLLSRI
  SASFADQGTRELUPFYQYSAGGIGSLRGFAYGAPNALIYRTQCPDSDYCLVSSDV1
  GNMATYFSLVLPFFAFVQKNSVRVTSLFVDAASVWNRKAEADKAKFALNVPD
  YSDPSVRASAGVALQWQSPIGPLVFSYAKPLKKYQCGDETEQFQFSIGGTF"
  BASE COUNT 3056 a 1752 c 2395 g 2984 t
  ORIGIN

alignment_scores
  Quality: 1054.00      Length: 829
  Ratio: 1.948          Gaps: 17
  Percent Similarity: 65.259  Percent Identity: 31.242

alignment_block
  US-09-701-711-2 x AE006235
  Align seg 1/1 to: AE006235 from: 1 to: 10187
```

```

9 pheGlnValSerAlaMetThrMetalValMetMetValMetSerThrHi 25
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
  7722 TTCACGATGAAAAAATTTTAAATTCGAGCTTATTATTGGTTCAACCAC 7771

25 salAGlnAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuG 42
  ||| ||||| : : : : : : : : : : : : : : : : : : : : : :
  7772 TGCATTTGCTGCGCCGCTTTGTAGTAAAGACATTCGTGTGACGGTGTTC 7821

42 InArgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGly 58
  ||| ||| : : : : : : : : : : : : : : : : : : : : : :
  7822 AGCAGGTACAGAAGAGTATAGCTACACTCTCTGTCGTGTGGG 7871

59 GlnValValSerGluAsnGlnLeuAlaAspGlyValValAlaLeuTyrAl 75
  ||| : : : : : : : : : : : : : : : : : : : : : :
  7872 CAGCGAGCAACAGATAACGATATGCTAATGCTGACGAAATATTATTCCT 7921

75 aThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleI 92
  : : : : : : : : : : : : : : : : : : : : : :
  7922 GAGTGGCAATATGATGATGTGAAAGCAAGTCGCGAAGGAATACATTAG 7971

92 leTyrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGly 108
  : : ||| : : : : : : : : : : : : : : : : : : : : : :
  7972 TTGTGACGTCATGCTTAACCTGTATTTCAAACGTCCTGATTGACGGT 8021

109 AsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaG 125
  ||||| : : : : : : : : : : : : : : : : : : : : : :
  8022 AATAAATCGATTCCTGATGAGCAATTAACAAACCTTAGATCGGAATGG 8071

125 yLeuAlaValGlyGlnProLysGlnAlaThrValGlnMetIleGluT 142
  : : : : : : : : : : : : : : : : : : : : : :
  8072 CTTTAAAGTCGGTGTGATTAACCCGCTAAATTAGAAGAAATTCGCGA 8121

142 hrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIle 158
  : : : : : : : : : : : : : : : : : : : : : :
  8122 AAGGATGTGCAACACATACATAGTGTGCGCTATTAATCGGAAGGTA 8171

159 ThrValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetH 175
  : : : : : : : : : : : : : : : : : : : : : :
  8172 GAGGTCATCGTCACTACATACCAATAATAGCGGAAATTAATAATCA 8221

175 rPheAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyA 192
  : : : : : : : : : : : : : : : : : : : : : :
  8222 AATTAATGAAGATGATGTGCGCTATTTAAAGAAATATTATTTGAAGGTA 8271

192 snGlnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysasp 208
  ||||| : : : : : : : : : : : : : : : : : : : : : :
  8272 ATCAAGCATTTAGCAGCAGTAATTAGAAGATCAATGCAACTTCAACA 8321

209 AsnLysIleAsnProLysLysAlaAspArgTyrThrGlnGluLysLe 225
  : : : : : : : : : : : : : : : : : : : : : :
  8322 GATGCATGCTGGAATTTGTTT...GGTAACAAATTTGATCAACCCCAAT 8368

225 uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV 242
  : : : : : : : : : : : : : : : : : : : : : :
  8369 CAATAAAGATTTAGAAACCTTACGTAGCTATTATTATGATCGTGTTCG 8418

242 alArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258
  : : : : : : : : : : : : : : : : : : : : : :
  8419 CGCAATTCCAAATTTTACATCTACTGATATCAAAATTAAGTATGATAAAAA 8468

259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheG 275
  : : : : : : : : : : : : : : : : : : : : : :
  8469 GAAAGCGCTGTCATATTATAAGTGAAGAAGGTGACTTATATACAGTGA 8518

275 yGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluA 292
  : : : : : : : : : : : : : : : : : : : : : :
  8519 AAGCGCGGTATTCTGGGGGATGTGGTGCGATGTCACGACGAACTTGCTC 8568

292 laLeuLeuLysPheLysAlaGluGluGly...PheSerGlnAlaMetLeu 307
  : : : : : : : : : : : : : : : : : : : : : :
  8569 CGATTTTAGATGCGATTCACTAAATGCTCTTTCCGTCGCGCAACGTA 8618

308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspGlyTyrTy 324
```

```

8619 TTGGAAGTTGAACAACGCAATTAATCGAAGTTAGGTGAAGAGGTTATGC 8668
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
8669 GACTGGCGCAAGTCATGTTCCACCGCACATTTTGACGACACAGATAAACGA 8718
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
341 aAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
8719 TTTGGTTAGATTTTATTGTTGAAGCAGGCAAAAGTTATACGGTTGCCAA 8768
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValIleuArgArgG 374
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
8769 ATTCTGTTTGAAGCAATCAACAGTAGTCAGATAGCACCTTACGTCAGGA 8818
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
8819 AATGCCGTCAACAAGAGGGCTTTGGCTATCTCCGCGAGTTGGTTGAGTTAG 8868
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
391 eArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
8869 GTAAATTACGTTTATAGTCGTACGGGATCTTTGAGAGTGTAGAAACCAA 8918
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
8919 ACAGAAGCTATCCAGGTTCT...GATCAAGTCGATGTGATTATAAAGT 8965
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
424 lGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnS 441
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
8966 CAAAGAGCGTAATACGGGTAGCAATTAACCTTGGTATGGTTATGGTACAG 9015
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
9016 AAAGTGGTTGACCTACCAAGCCAGTATTAACACAGGATAACTTCTTAGGA 9065
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
9066 ATGGGATCTCTATTAGTTAGTGGGACGCGTAATGACTACGCTACTAC 9115
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
9116 AATCAATCTGGTTATAGAGCGGTACTTTACCAAGATGGTGTG... 9161
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
491 lnSerLeuSerGlyTyrTyrArgLysThrLysTyrAspAsnLysAsnIle 507
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
9162 ..AGCCTCGTGCGCAATGTTTCTTTGAAGATATGATAGTTCCTCAAAAGT 9209
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
508 SerAsn.....TyrValLeuAspSerTyrGlyGlySerLeuSerTy 521
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
9210 AATACCTCTGCAGCCTATGACGCGACTAGCTATGCTGGTAAATTGACACT 9259
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
521 rGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaA 538
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
9260 AGCCTTCGGTGAATGAGTAATCACTATATATCTTGGTGTGGGCTATA 9309
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
538 sPasnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnVal... 553
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
9310 CGTATAATAAAGT... ..AAGATATACGCA 9335
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
554 LysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnGlyIle 570
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
9336 CCGGAATAAATCGTATTATATCTCAATCAATGAAATATAATGATTC 9385
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
570 eProAspPheLys...HisAspTyrThrThrTyrAsnAlaIleLeuGlyT 586
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
9386 TTGGACCTTTAAATCCAGCAT... ..TTGATTTGTCCTTTTGGTT 9426
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
586 rPasnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSer 602
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
9427 GGAATTATACAGCTCTACCGGTGGCTATTTCCTCAACCAAGGGGTACGT 9476
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
603 HisSerValAsp.....LeuThrValGlyPheGlyAspLysThrHisG 617
      ::::: ||||| ||||| ||||| ||||| ::::: :::::

```

```

9477 GCCAATATTGGTGGACGAGTGACCATTCGGGGCTCAGACAATAAATATTA 9526
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
617 nLysValValTyrGlnGlyAsnIleTyrArgProphe..... 629
      ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
9527 TAAACTCAATGCAGAACACAGGGTTCTATCCGTTAGATCGTGAACATG 9576
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
630 .....IleLysLysSerValLeuArgGlyTyrAla 639
      ||||| ::::: ||||| ::::: :::::
9577 GTTGGGTACTTTCAAGCCGTATTAGTGCCTCTTTTGTGATGATTTGGC 9626
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
640 LysLeuGlyTyrGlyAsnAsnLeuProPheTyrGluAsnPheTyrAlaG 656
      ||||| ::::: ||||| ||||| ||||| ::::: |||||
9627 .....GGTAAGCGTTTGGCGTTCTATCAATATTATAGCGCAGG 9664
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
656 yGlyTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgS 673
      ||||| ::::: ||||| ||||| ||||| ::::: :::::
9665 CGGTATCGGGAGTTTACGTGGCTTTGGCTATGTTGGTATGGACCAATG 9714
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
673 erGlnAlaTyrLeuThrAlaArgGlyGlnGlnThrThrLeu..... 687
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
9715 CA...ATTATCGTACACGTCATGTCTCAGACGCTATTGTTAGTCAGT 9761
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
688 GlyGluValValGlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuI 704
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
9762 AGCGATGTGATTTGGGGGAATGCAATGTCACGCCAGTACCGAACTCAT 9811
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
704 eLeuProLeuProPheLysGlyAspTrp...IleAspGlnValArgProV 720
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
9812 TGCCCAACACCATTTGTCGAGATAAAACCAAACTCAGTAAGAATCT 9861
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
720 alilePheIleGluGlyGlyGlnValPheAspThrThrGlyMetAspLys 736
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
9862 CTTTGTGTTGGATGCGCAAGTGTGTGGAATACG..... 9896
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
737 GlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluG 753
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
9897 .....CGTTGGAAACGACAGGA 9913
      ::::: ||||| ||||| ||||| ||||| ::::: :::::
753 nAsnAlaLysAlaAla.....AsnArgProLeuLeuThrGlnAspLysG 768
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
9914 TAAAGCAAAATTTGCAAAATGGAATGTCGACGATCAGTGACCCAAGTC 9963
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
768 lnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleGly 784
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
9964 GCCTTCGTGCTTCAGCTGGGTGGCGCTTCAATGGCAATCGCCAATTGGA 10013
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
785 ProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAspG 801
      ||||| ::::: ||||| ||||| ||||| ::::: |||||
10014 CCGTTAGTGTCTCTTATGCAAACTCTTAAAAAATACCAAGCGGATGA 10063
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
801 nThrAspThrValGlnPheGlnIleGlySerValPhe 813
      ::::: ||||| ||||| ||||| ||||| ::::: |||||
10064 AATTGACCAATTCATTCAGCATTTGGTGGGACGTTTC 10100
      ::::: ||||| ||||| ||||| ||||| ::::: |||||

```

seq_name: gb_ba:U32773

seq_documentation_block:

LOCUS U32773

10151 bp DNA linear BCT 29-MAY-1998

DEFINITION Haemophilus influenzae Rd section 88 of 163 of the complete genome.

ACCESSION U32773 L42023

VERSION U32773.1 GI:1573932

KEYWORDS

SOURCE Haemophilus influenzae Rd.

ORGANISM Haemophilus influenzae Rd.

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

Haemophilus.

REFERENCE 1 (bases 1 to 10151)

AUTHORS

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A.,
 Kirkness, E.F., Kervatage, A.R., Bult, C.J., Tomb, J., Dougherty, B.A.,
 Merrick, J.M., McKenney, K., Sutton, G.G., Fitzhugh, W., Fields, C.A.,
 Gocayne, J.D., Scott, J.D., Shirley, R., Liu, L.I., Glodex, A.,
 Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E.,
 Cotton, M.D., Utterback, T., Hanna, M.C., Nguyen, D.T., Saudek, D.M.,

Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L., Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M., Smith,H.O. and Venter,J.C.
Whole-genome random sequencing and assembly of Haemophilus influenzae Rd
Science 269 (5223), 496-512 (1995)
2 (bases 1 to 10151)
Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S., Borodovsky,M., Rudd,K.E. and Koonin,E.V.
Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli
Curr. Biol. 6 (3), 279-291 (1996)
3 (bases 1 to 10151)
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
Direct Submission
Submitted (25-JUL-1995) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
4 (bases 1 to 10151)
White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
Direct Submission
Submitted (27-SEP-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
The H. influenzae sequence has been updated by R. Fleischmann. New database matches have been assigned, product names have been improved, and a number of frame shifts have been corrected. We gratefully acknowledge the work of Tatusov et. al. We have incorporated their annotation into the /notes fields of the corresponding H. influenzae genes
5 (bases 1 to 10151)
White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D., Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
Direct Submission
Submitted (28-MAY-1998) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
The whole genome was shifted by 588 nucleotides for a new start
On Sep 30, 1996 this sequence version replaced gi:1221647.
Location/Qualifiers
1..10151
/organism="Haemophilus influenzae Rd"
/db_xref="taxon:71421"
81..845
/gene="HI0912"
81..845
/gene="HI0912"
/note="hypothetical protein; identified by GeneMark; putative"
/codon_start=1
/transl_table=11
/product="H. influenzae predicted coding region HI0912"
/protein_id="AAC22579.1"
/db_xref="GI:1573933"
/translation="MNKISVYDKDNFFELYQKLRANPISLNEIIEKPTWLSLLPNLK GKLLDCCGTGGHLQYLERGAKVIGTDLSEKMLEQAKDKQKQFSGRFSLYHL PIELAEPSHEDFTTSFAPHYIENFTLLSEIKHKLSSNGTLIFSQSHPTTCHK EGERWEKNDKQOVAYLNHYRREGKRNWFKQFQYHRTTATINNLHARFQIE QMEEPMLADQPQWVNEFKLSHRPPLLFIKARKVEK"
990..1745
/gene="HI0913"
990..1745
/gene="HI0913"
/note="similar to GB:D26562 SP:P02351 GB:V00343 PID:42842 PID:473824 percent identity: 82.16; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S2 (rps2)"
/protein_id="AAC22571.1"
/db_xref="GI:1573934"
/translation="MEAPQPOKENIMAOVSMRDMINAGVHFHGTWYNPOMKPFPIFG ARGVHINLEKTLPLFNEALIELTRASNNGKVLVFGTKRAAEAVQAALQQOQY VNHWRGGMLTNKVTQSKIKRLKDLTQSQDQFTFKLKKKALMRSEKLELSLG

GIKMGGLPDALFVIGADHEHIAVKEANNLIGIPVFAIVDTNSTPAGVDFVIPGNDAT
RAQLVYSAANAANKKEGRNEAQVAEELAADA"
1878..2729
/gene="HI0914"
1878..2729
/gene="HI0914"
/note="similar to GB:D26562 SP:P02997 GB:D13334 GB:V00343
PID:1208943 percent identity: 71.43; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="elongation factor Ts (tsf)"
/protein_id="AAC22572.1"
/db_xref="GI:1573935"
/translation="MAETLSLVKELDRDTGAGMWECKKALVEANGDIELAIDNMRKS
GQAKAKKAGRAAEGVILARVENGFGVLYEMNCETDFVAKDAGFLGIANEYTFDAAA
NKGTTTIALQAQFEERKALVAKIGENNRIRVAYLDGVAIQIHLGAKIGLVVAGEG
SADELKAMHVAASKPEFVNEDVSAEVEHEHQIQIDIAINSGKKEKEAEKAVEGR
MKRPTGEVSLTGOAFVMDPSVSGDFLKSVTNSVSNFIRLEVGEKIEKEDFAAEVA
KITGGNA"
complement(2909..3934)
/gene="HI0915"
complement(2909..3934)
/gene="HI0915"
/note="similar to GB:X54797 SP:P21645 PID:1208950
PID:41470 GB:U00096 percent identity: 65.58; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="UDP-3-O-(3-hydroxymyristoyl)-glucosamine
N-acetyltransferase (lpxB)"
/protein_id="AAC22573.1"
/db_xref="GI:1573936"
/translation="MQKXSLOELATOIGATVRGNTDVVVENTAPLDKASQNLTFIS
NWFRVLKDSKAGILVSEEDVEHCPESNLLIVKDPYVAYAILAQMSTPKAAOG
IAKSAVFDGVLGNCISGANAVIEGVVLDGNDVIIGANCFVGKNTKIGSGTOLWAN
VTYVHNVEIGANGLIQSGTVIGSDFGFYANDRGWRKIPQVGQVILIGNNVEIGANTCI
DRGALDATTIEDNVIDNLQIAHNHVGITGAVAGVIMAGSLTVGRYCLIGASVI
NGHMEICDKVTITGMGMVRCPIPEPGVSSGIPLOTNKEWRKTAALTGLDIGNKRLK
ALEKKIS"
complement(3947..4540)
/gene="HI0916"
complement(3947..4540)
/gene="HI0916"
/note="similar to PID:1935023 percent identity: 37.58;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein, putative"
/protein_id="AAC22574.1"
/db_xref="GI:1573937"
/translation="MKNIAKVTALALGIALASGYASAEKIAFINAGYIFQHPDROA
VADKDAEFKPAEKLAASKKEVDKTAARKKVEAKVALEKADAPRLROADIQKROO
EINKLGAEDAELOKLQEQDKKVOEQAKNEKQAEERCKLLDSIQATNNLAKAG
YTVLDANSIVFAVEGKDITEEVLKSPASEKAQEKK"
complement(4648..7035)
/gene="HI0917"
complement(4648..7035)
/gene="HI0917"
/note="similar to GB:U13961 PID:1573938 PID:2231192
PID:2231195 percent identity: 98.74; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="protective surface antigen D15"
/protein_id="AAC22575.1"
/db_xref="GI:1573938"
/translation="MKKLLIASLLFGTTTTFVAFVAKDIRVDGVGDLEEQIQRASL
IPVARGQRTDNDVANIIVRSFLVSGRFDVKAHQEGDVLVSVVAKSISDVKIGKNSI
IPTPEALKQNDLNGFKVGDVILIREKLNFAKSVKHEVSVGRYNAIVPIVNTLPNNR
AEILIOYNEDDKAKLASLTFKGNESSVSSLOEQMLQDPDSWKLWNKPEGAQPEKD
LOSTRDYLANGYAKAQTITQDNLDEKTKVNTVIDNSELQVLDLSARLIGLGM
SAELEPLLSALHLNDFRRSDIADVENAIKAKLGERGSGSATVNSVDFDANKTLAI


```
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1690 GGAAGTAATGTTACTTTAGTTTCCCTCTAAATGAAATAACTCTATTA 1739
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1740 TGTAGGATAGGTATACCTATATAATAATAGTAACCTTGTCTAGAAAT 1789
544 LyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1790 ATAACCGTAATTTATATATCAATCAATGAAA..... 1821
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1822 .....TTTAAAGGTAATGCCATTAACAAACAAATGACTTTGATT 1859
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1860 TTCT.....TTTGGTTGGAACATATAACAGCCCTTAATAGAG 1894
594 roValPheProThrGlnGlyMetSerHisSerValasp.....Leuthr 608
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1895 GCTATTTCCCACTAAAGGGTTAAGCAAGTCTTGGTGGACGAGTTACT 1944
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnII 625
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1945 ATTCCAGGTCTTGATAACAATACTACAAACTAAGTGCAGATGTACAGGG 1994
625 eTyrArgProPhe.....IleLysLysSerValL 635
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1995 TTCTTACCATTAGACAGAGATCACTCTGGTGTGTATCTGCAAAAGCAT 2044
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2045 CTGCAGGATATGCAAAAT...GGTTTGGAAACAACGGTTTACCGTCTCAT 2091
651 GluAsnPheTyrAlaGlyClyTyrClySerValArgGlyTyrAspGlnSe 667
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2092 CAAACTTATACAGCGGGTGGCATCGTTCATTACGTGTTTGTCTTATGG 2141
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2142 TAGTATTCGACCTACGCA.....ATTATGCCGAATATGTTAATG 2182
684 InThrThrLeuGly.....GluValValGlyGly 693
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2183 GTAGTGGTACTGGTACTTTTAAAGATAAGTCTCTGATGTGATGGTGT 2232
694 AsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheLy 710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2233 AATGCAATCGCTACAGCTAGCGCAGAGTTAATTTGCAACTCCATTGT 2282
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGly 726
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2283 GAGCGATAAGACCAAAATACGGTCCGAACCTCTTATTGTTGATGCGG 2332
726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2333 CAAGTGTGTAATACTAAATGAATCAGATAAAATGGATTAGAGAGC 2382
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaI 758
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2383 GATGTATTAATAA..... 2394
758 aAsnArgProLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2395 .AGATTCCTGATTATGCAATCAAGCCGATTTCGCCCTCTACAGGTG 2443
775 alGlyAlaThrTyrThrProIleGlyProLeuSerIleSerTyrAla 791
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2444 TCGGATTCCAATGCAACTCTTATGGCCATTCGTTATCTCTATGCC 2493
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
```

```
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2494 AAACCAATATAAAATATCAAAATGATGATGTCGAACAGCTTCCAATTAG 2543
808 nIleGlySerValPhe 813
|||||:|||||:|||||:|||||:|||||:|||||:
2544 TATTGGAGGTCTTCTTC 2559
seq_name: gb_ba:HIU60833
seq_documentation_block:
LOCUS HIU60833 2776 bp DNA linear BCT 02-JUL-1997
DEFINITION Haemophilus influenzae D15 (d15) gene, complete cds.
ACCESSION U60833
VERSION U60833.1 GI:2231194
KEYWORDS
SOURCE
ORGANISM
Haemophilus influenzae.
Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
Haemophilus.
REFERENCE
1 (bases 1 to 2776)
AUTHORS
Loosmore,S.M., Yang,Y., Coleman,D.C., Shortreed,J.M., England,D.M.
and Klein,M.H.
TITLE
The gene encoding outer membrane protein D15 is highly conserved
amongst Haemophilus influenzae species and may be part of an operon
Unpublished
REFERENCE
2 (bases 1 to 2776)
AUTHORS
Loosmore,S.M.
DIRECT SUBMISSION
TITLE
Submitted (13-JUN-1996) Loosmore S. M., Connaught Laboratories
Limited, Molecular Genetics Research, 1755 Steeles Avenue W., North
York, Ontario M2R 3T4, Canada
FEATURES
Location/Qualifiers
source
1..2776
/organism="Haemophilus influenzae"
/isolate="Minna"
/species="Homo sapiens"
/db_xref="taxon:727"
/clone="DS-691-1-5"
328..2721
/gene="d15"
328..2721
/gene="d15"
/codon_start=1
/transl_table=11
/product="D15"
/protein_id="AAB61976.1"
/db_xref="GI:2231195"
/translation="MKLLIASLLFGTTTTFVFAAPVFAKDIRVDVGQDLEQIRASI.
PVRAGQRTDNDVANIVRSIFVSGRFDDVKAHQEDVLVSVVVAKSIIISDVKIRGNSV
IPTALKQNDLNGFKVGDVLIIEKLNFEAKSVKHEHYASYGRYNAVTEPTVNTLNNR
AETLIQINEDDKAKLASLTFKGNESVSSSTLQEMELQPDSPWKLGNKFGAOFKED
LQSIDRYLNGYAKAQTITKTDVQNDKTKVNTIDVNEGLOYDLRSARIIGNLGM
SAEPLLSALHLNDTFRSDIADVENAIKALGERGYSATVNSVPDFDDANKTLAI
TLVDAGRLTLVRLRFEPTVSADSLRQEMRQQTWNSQLVELGKIRLDTGFF
ETVENRIDPINGSNDVDVYVYKVKERTGSINFGIYGTEGSIYSQASVKODNPLCTG
AAVSIAGTKNDYGVNLYGTEPTFKDGVSLGNGVFFENYDMSKSDTSNKRVTYG
SNVTLPFPVNNNSYVGLGHTYKINSFALEYNRLYIQSMFKGNGIKTNDDFSF
GNVNSLNRGYFPTKGVKASLGRVTPGSDNKKYKLSADVQGFYPLDRHLAWVSAS
ASAGYANGFGNKLPPFYQTYTAGTIGSLRGFAYSIGPNALYAEYNGSGTGTFFKIS
SDVIGGNATASAEILVTPFVSDSKSNTVRSLSFVDAASVWNTWKMSDKNGLESVD
LKRLPDYKGKSSRIRASTGPGFOWSPIGLTVFSYAKPIKKIYENDVDEQFSGGSF"
BASE COUNT 888 a 435 c 588 g 865 t
ORIGIN
alignment_scores:
Quality: 1052.50 Length: 822
Ratio: 1.956 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630
alignment_block:
US-09-701-711-2 x HIU60833
```

Align seg 1/1 to: HU60833 from: 1 to: 2776

```
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   ::::: ::::: ||||| ::::: ||||| ::::: ||||| :::::
367 ACAACGACACTGTTGTCGCCACACTTTTGGCAAAAGATATTGCGTGGGA 416
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
417 TGGTGTCAAGTGACTTAGACACACAATCCGAGCAAGTTTACCTGTTTC 466
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
56 rGluGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   || ||||| ||||| ::::: ||||| ::::: ||||| ::::: |||||
467 GTCCGGTGCAGCGTGTGACTGACATGACAAATGATGGCTAAATATTG 516
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
73 LeuTyAlaThrGlyAsnPheSerAspValGlnValTyHisGlnGluGl 89
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
517 TTATTTCGTAAGTGGTGGATTCGATGATGGAAGCG...CATCAAGAAGG 563
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
89 yArgIle...IleTyGlnValThrGluArgProLeuIleAlaGluIleA 105
   | ::::: ::::: ||||| ::::: ||||| ::::: ||||| :::::
564 CGATGTGCTGTGTTGAGGTGTCGCTAAATCGATCATTTTCAGATGTTA 613
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
   ::::: ||||| ||||| ::::: ||||| ::::: ||||| ::::: |||||
614 AAATCAAGGTAACCTCTGTATTCCACTGAAGCACTTAAACCAAACTTA 663
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
122 LysAsnAlaGlyLeuAlaValGlnProLeuLysGlnAlaThrValGl 138
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
664 GATGCTACGGGTGTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTA 713
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
138 nMetIleGluThrGluLeuThrAsnGlnTyIleSerGlnGlyTyTyra 155
   | ::::: ::::: ||||| ::::: ||||| ::::: ||||| :::::
714 TCAATTTGCCAAAGTGTAAAGAGACCACTATCCAACTGTAGTCGCTATA 763
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
764 ACGCAACAGTTGAACCTATTTGTCACTACCTACCAAAATAATCGCGTGA 813
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIle 188
   | ::::: ::::: ||||| ::::: ||||| ::::: ||||| :::::
814 ATTTTAATTCAAATCAATGAAGATGATAAAGCAAAATGGCATCATTAAC 863
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
188 nIleLeGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeu 205
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
864 TTTCAAGGGAACGAATCTGTTAGTAGCAGTACATTACAAGAACAAATGG 913
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
205 laileLysAsnLysIleAsnProLeuSerLysAlaAspArgTyThr 221
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
914 AATTACAACCTGATTCTTGGTGGAAATTA...TGGGGAATAAATTTGAA 960
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyLeuAs 238
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
961 GGTGCGCAATTCGAGAAAGATTGGCAGTCAATTCGTATTATTATTAA 1010
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsn 255
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1011 TAATGGCTATGCCAAGACACAAATTAATAAACGGATGTTACGTAATG 1060
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1061 ATGAAAAACAAAGATTAATGTAACCATTTGATGTAATGAAGTTTACAG 1110
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyThrGlnAl 288
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1111 TATGACCTTCGTAGTCAGCAATATTAGTAAATCTGGGAGGTATGTCTGC 1160
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
288 aGluLeuAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1161 CGAGCTTGAACCTTACTTCTTACGATTTACATTTAAATGATACTTTCCGCC 1210
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
304 lnaAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
1211 GTAGTGATATTCCAGATGTAGAAAAATGCAATTTAAAGCAAAACTTGA 1260
```

```
321 AspGlyTyTyTyAlaGlnIleArgProValThrArgIleAsnAspGl 337
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1261 CGCGGTTACGGTAGCGCAACCGTAAATTCAGTACCTGATTTTGTATGATGC 1310
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
337 uSerArgThrValAspValGluTyTyTyIleAspProValHisProVal 354
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
1311 AAATAAACATTAGCGATAACCCCTTGTGTGATGCTGGACGACGTTAA 1360
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
354 yrvAlaArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1361 CTGTTCCCAACTTCGCTTTGAAGAAATACCGTTTCTGCTGATAGCACT 1410
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1411 TTACGTGAGGAATGCGCAACGAAGAACTTGGTATAAATTCACAAT 1460
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
1461 AGTTGAGTTAGGAAAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1506
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1507 ..ACAGTCGAAACCGAATTTGATTCCTTATCAATGGTAGTAATGATGAAGT 1554
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1555 GATGCTGATATAAAGTCAAGAACGTACACGGGTAGTATCAACCTTTGG 1604
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
435 aAlaGlyTySerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1605 TATTGTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1654
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
452 lnaAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1655 AAGATAAATTTCTGGGACACGGGCGCAGTAGTAGTATGCTGTAGCAAA 1704
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
469 SerGluThrArgGluValTySerLeuGlyMetThrAsnProTyPheTh 485
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1705 ATGATTATGTTACGAGTGTCAATTTGGTATTACCGAGCCCTATTTTAC 1754
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
485 rValAsnGlyValSerGlnSerLeuSerGlyTyTyTyArgLysThrLys 502
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1755 TAAAGATGCTGA....AGTCTTGGTGAATGTTTCTTTGAAAAACT 1798
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
502 yvAspAsnLys.....AsnIleSerAsnTyTyTyValLeuAspSerTy 515
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1799 AGGATAAATCTTAAAGTGTATACATCTCTTAACCTATAAGCGTACGACTTAC 1848
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
516 GlyGlySerLeuSerTyGlyTyProIleAspGluAsnGlnArgIleSe 532
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1849 GGAAGTAATGTTACTTTAGTTCCTCTGTAATGAAATAAATCTCTTATTA 1898
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1899 TGTAGGATTAGTGTATACCTATAATAAATAATTTAGTACTTTGCTCTAGAAT 1948
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
1949 ATAACCGTAATTTATATATTCATCAATCAATGAAA..... 1980
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
1981 .....TTTAAAGGTAATGCAATTAACCAAAATGACTTTTCATTT 2018
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
577 rThrThrTyAsnAlaIleLeuGlyTyTrpAsnTySerSerLeuAspArg 594
   ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
2019 TTCT.....TTTGGTTGGAACATATAACAGCCTTAATAGAG 2053
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
   ||||| ::::: ||||| ::::: ||||| ::::: ||||| :::::
2054 GCTATTTCCCACTAAAGGGTTAAAGCAAGTCTTGGTGGACGAGTACT 2103
```

609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
 2104 ATCCAGGTTCTGATACAAATACTACAACTAAGTCAGATGACAGG 2153
 625 eTyrArgProPhe.....IleLysLysSerValL 635
 2154 TTCTACCATACAGAGATACCTCTGGGTTGTATCTCAAAAGCAT 2203
 635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
 2204 CTGAGGATATGCAAAAT...GGTTTGGAAACAAGCGTTTACCGTCTAT 2250
 651 GluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnse 667
 2251 CAACCTTATACAGCGGTGGCATCGGTTCACTACGTTGTTTGGTTATGG 2300
 667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
 2301 TAGTATTGGACCTACGCA.....ATTATGCCGAATATGTAATG 2341
 684 InThrThrLeuGly.....GluValValGlyGly 693
 2342 GTAGTGGTACTGCTACTTTTAAAGATAGTCTCGATGATGTTGGT 2391
 694 AsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPhe 710
 2392 AATGCAATCGTACAGTACGAGGAGTAAATGTCGCAACTCCATTGT 2441
 710 sGlyAspTrpIle...AspGlnValArgProValIlePheLeuGlyG 726
 2442 GAGCGATAAGACCAATAACGTCGCAACCTCTTATTGTTGATGGG 2491
 726 yGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
 2492 CAAGTGTGGAATACTAAATGGAATCAGATAAAATGGATTAGAGAGC 2541
 742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
 2542 GATGATTATAAA..... 2553
 758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGly 775
 2554 AGATTGCTGATTATGCAATCAAGCGTATTCGCGCTCTACAGGTG 2602
 775 aGlyAlaThrThrTyrThrProIleGlyProLeuSerIleSerTyrAla 791
 2603 TCGGATTCAATGGCAATCTCTATTGGCCATTGGTATTCTTATGCC 2652
 792 LysProLeuAsnLysGlnAsnAspGlnThrAspThrValGlnPheG 808
 2653 AAACCAATATAAAATATGAAATGATGATGTCGAACAGTTCCAATTTAG 2702
 808 nIleGlySerValPhe 813
 2703 TATTGGAGGTTCTTTC 2718
 seq_name: gb_ba:HIU60834

seq_documentation_block:
 LOCUS HIU60834 2823 bp DNA linear BCT 02-JUL-1997
 DEFINITION Haemophilus influenzae D15 (d15) gene, complete cds.
 ACCESSION U60834
 VERSION U60834.1 GI:2231196
 KEYWORDS
 SOURCE Haemophilus influenzae.
 ORGANISM Haemophilus influenzae.
 Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 Haemophilus.
 REFERENCE
 1 (bases 1 to 2823)
 AUTHORS Loomore,S.M., Yang,Y., Coleman,D.C., Shortreed,J.M., England,D.M.
 TITLE The gene encoding outer membrane protein D15 is highly conserved
 amongst Haemophilus influenzae species and may be part of an operon

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2823)
 AUTHORS Loomore,S.M.
 TITLE Direct Submission
 JOURNAL Submitted (13-JUN-1996) Loomore S. M., Connaught Laboratories
 Limited, Molecular Genetics Research, 1755 Steeles Avenue W., North
 York, Ontario M2R 3T4, Canada
 FEATURES
 Location/Qualifiers
 1..2823
 /organism="Haemophilus influenzae"
 /isolate="PAK 12085"
 /specific_host="Homo sapiens"
 /db_xref="taxon:727"
 /clone="JB-1042-9-4"
 390..2771
 /gene="d15"
 390..2771
 /gene="d15"
 /codon_start=1
 /transl_table=11
 /product="D15"
 /protein_id="AAB61977.1"
 /db_xref="GI:2231197"
 /translation="MKLLIASLLFGTTTFAAPFAVKDIRVDGVQGLEEQOIRASL
 PVKAGQRTDNDVANIIVRSLSVGRFDDVRAHQEGDVLVSVVAKSLISDVKIKGNSV
 IPTALKQNDLAKFKVGVGLIREKLEFAKSYKHEYASGRYNATVEPINTLPPNR
 AEILIOINEDDKAKLASLTFKNGSVSSSTLOQMELOPDQSMWKLGNKFEQAPEKD
 LOAIRDYILNNGYAKAQITKTDVQNDKTKVNTIDVNEGLOYDLRSARLIGNLGM
 SAELEPLLSALHLNDFRSDIADVENAIKAKLGERGYNTVNSVDFDDANKTLAI
 TFVDAGRRUTVRQLREFGNTVSADSLRQEMRQOEGTWNVSOLVELGKIRLDRTGFF
 ETVENRIDPINGSNDVYVYKVKERTIGSINFGIGTGESGISYQTSIKQDNPLFTG
 AAVSIAGTKNDYGSVNLGYTEPYFTKDGVSGLGNIFENFENYDMSKSDTSNKRRTTYG
 SNVTLGPPNNENSVYVGLGHTYNTKISNFALEYNNRNLIQSMFKKNGIKTNDPDSF
 GWVNSLNRGYFTKVGKSLGGRVTIPGSDNKYIKLSADVQGFYPLDRHRWVVSAS
 ASAGYANGCNKRLPEYQVYTAGSISLGRFAYGSGPNAIYAEHNGTGNKISSDVI
 GGNATTTASRIEPIYFPFVSDKQNTVTSILFVDAASVWNTKNSDKNGLESKVLKDL
 PDYKSSIRASTGVGFQWQSPGLVFTSAKPIKKTENDDDVEQFQFSIGGSF"
 BASE COUNT 918 a 442 c 588 g 875 t
 ORIGIN

alignment_scores:
 Quality: 1052.50 Length: 818
 Ratio: 1.953 Gaps: 21
 Percent Similarity: 65.892 Percent Identity: 31.785
 alignment_block:
 US-09-701-711-2 x HIU60834
 Align seg 1/1 to: HIU60834 from: 1 to: 2823
 23 SerThrHisAlaGlnAlaAlaAspPheMetaAlaAsnAspIleThrIleTh 39
 429 ACAACGACGTGTGTTGCCGACCTTTTGGCAAAAGATATTCTGTGTGA 478
 39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPhe 56
 479 TGGTGTTCAGGTGACTTAGAACACAAATCCGACGAGTTTACCTGTTTC 528
 56 rLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
 529 GTGCTGCTACGCTGTGACTGACATGATGCTGCTAATATTGTCGCTCT 578
 73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
 579 TTATTTCGTAAGTGTGCTGATTCGATGATGTAAGCG...CATCAAGAAG 625
 89 yargIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIle 105
 626 CGATGCTGTGTTGTTAGCGTGTGCTGCTAAATCGATTCATTCAGATGTTA 675
 105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGlyLeu 121

[illegible][illegible]

DEFINITION Sequence 1 from patent US 6083743.

ACCESSION AR102148

VERSION AR102148.1 GI:12812946

KEYWORDS

SOURCE

ORGANISM

Unknown.

Unclassified.

REFERENCE 1 (bases 1 to 2949)

AUTHORS Chong, P., Thomas, W., Yang, Y. Ping, Loosmore, S., Sia, D. Yuan Charles

TITLE Haemophilus outer membrane protein

JOURNAL Patent: US 6083743-A 1 04-JUL-2000;

FEATURES Location/Qualifiers

1..2949

source

BASE COUNT 984 a 475 c 613 g 877 t

ORIGIN

alignment_scores:

Quality: 1052.50 Length: 822

Ratio: 1.956 Gaps: 20

Percent Similarity: 65.450 Percent Identity: 31.630

alignment_block:

US-09-701-711-2 x AR102148 ..

Align seg 1/1 to: AR102148 from: 1 to: 2949

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39

114 ACAACGACTGTTGGCGACCTTTTGGCAAAAGATATCGTGTGGA 163

39 rclyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56

164 TGGTGTCAAGGTGACTTAGAACAAATCCGACGAGTTTACCTGTTC 213

56 rgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72

214 GTGCGGTGACGCGTGTGACTGACATGATGTGGCTAATATTGTCGCTCT 263

73 LeuTyrrAlaThrGlyAsnPheSerAspValGlnValTyrrHisGlnGluG 89

264 TTATTCGTAGTGTGATCGATCGATGTGAAGCG...CATCAAGAGG 310

89 yArgIle...IleTyrrGlnValThrGluArgProLeuIleAlaGluIleA 105

311 CGATGTGCTTCTGTAGCGTTGTGGCTAAATCGATCATTTTCAGATGTTA 360

105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121

361 AAATCAAGGTAACCTCTGTATTCCCACTGAAGCACTTAAACAAACTTA 410

122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138

411 GATCCTAACGGGTTAAAGTTGGCGATGTTTAAATTCGAGAAATTA 460

138 nMetIleGluThrGluLeuThrAsnGlnTyrrIleSerGlnGlyTyrrA 155

461 TGAATTTGCCAAAAGTGTAAAGAGCACTATGCAAGTGTAGTGCCTATA 510

155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171

511 ACGCAACAGTTGAACCTATTGTCATACGCTACCAAAATATCGCGCTGA 560

172 LeuAspMetThrPheAlaGlyLysProAlaArgValValAspIleAs 188

561 ATTTTAAATCAATCATGAGATGATAAGCAAAATTTGGCATCATTAAC 610

188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205

611 TTTCAGGGGAACCAATCTGTGTAGTACGATGATATTACAAGCAAAATGG 660

205 laileLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrrThr 221

661 AATTAACAACCTGATTCCTGGTGAATA...TGGGGAATAAATTTGAA 707

222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrrLeuAs 238

708 GGTGGCAATTCGAGAAAGATTTCAGTCAATTCGTGATTATTATTAA 757

238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255

758 TAATGGCTATGCCAAAGACAAATTTACTAAACGGATGTTTCAGCTAAATG 807

255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271

808 ATGAAAAACAAGAGTTAATGTAAACCATTTGATGATAATGAAGTTTACAG 857

272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrrThrGlnAl 288

858 TATGACCTTCGTAGTCACGCAATTATAGTAATCTGGGAGGTATGCTGTC 907

288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304

908 CGAGCTTGACCTTTTACTTTTCAGCATTTACATTTAAATGATACTTCCGCC 957

304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320

958 GTAGTGATATTGCAGATGTAGAAAAATGCAATTAAGCAAAACTTGGAGAA 1007

321 AspGlyTyrrTyrrAlaGlnIleArgProValThrArgIleAsnAspG 337

1008 CGCGTTACGTTAGCGCAACCGTAAATTCAGTACCTGATTTTGTGATGTC 1057

337 uSerArgThrValAspValGluTyrrTyrrIleAspProValHisProVal 354

1058 AAATAAACAATTCAGCAATACCTTGTGTGATGCTGGACGACGTTTAA 1107

354 yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370

1108 CTGTCGCCAACTTCGCTTTGAAGGAAATACCGTTTCTGCTGATAGCACT 1157

371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnly 387

1158 TTACGTCAAGAAATGCGCCACAGAGAGAACTTGGTATATATTCACAAT 1207

387 sileGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404

1208 AGTTGAGTTAGGAAAAATTCCTTAGATCGTACAGGTTTCTCGAA... 1253

404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418

1254 ..ACAGTCGAAAACCGAATTCATCTATCAATGGTAGTAAATGATGAAGT 1301

419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435

1302 GATGCTGATATAAAGTCAAGACGTAACACGGGTAGTATCAACTTTGG 1351

435 aAlaGlyTyrrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452

1352 TATTGGTTACGTTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1401

452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468

1402 AAGATAATTTCTGGGAACAGGGCGGCGAGTAAGTATATAGCTGGTACGAAA 1451

469 SerGluThrArgGluValTyrrSerLeuGlyMetThrAsnProTyrrPheTh 485

1452 ATGATTTATGTTAGTGTCAATTTGGTTATACCGAGGCCCTATTATTAC 1501

485 nValAsnGlyValSerGlnSerLeuSerGlyTyrrTyrrArgLysThrLys 502

1502 TAAAGATGGTGTGA...AGTCTTGGTGGAAATGTTTCTTTGAAAAACT 1545

502 yrAspAsnLys.....AsnIleSerAsnTyrrValLeuAspSerTyrr 515

2350 TCGGATTCCAATGGCAATCTCTATTATGGGCCATTGGTATTCTCTTATGCC 2399

792 LysProLeuAsnLysGlnAsnAspGlnThrAspThrValGlnPheGI 808
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
 2400 AAACCAATTAAAAATATGAAATGATGATGTCGAACAGTTCCAATTAG 2449

808 ntleGlySerValPhe 813
 :|||||::: |||
 2450 TATTGGAGTCTTCTTC 2465

seq_name: gb_pat:A38694

seq_documentation_block:
 LOCUS A38694 2950 bp DNA linear
 DEFINITION Sequence 5 from Patent WO9412641.
 ACCESSION A38694
 VERSION A38694.1 GI:2295177
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 unclassified.
 REFERENCE 1 (bases 1 to 2950)
 AUTHORS Chong,P., Thomas,W., Yang,Y., Loosmore,S., Sia,D.Y.
 TITLE HAEMOPHILUS OUTER MEMBRANE PROTEIN
 JOURNAL Patent: WO 9412641-A 5 09-JUN-1994;
 COMMENT CONNAUGHT LAB (CA)
 Other publication CA 2149319 940609
 Other publication AU 5556594 940822
 Other publication JP 8502417T 960319.
 FEATURES
 source 1. 2950
 /organism="unidentified"
 /db_xref="taxon:32644"
 334..2727
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAA02390.1"
 /db_xref="GI:2295178"
 /translation="MKLLIASLLFGTTTTFVFAAPFVAKDIRV
 PVRAGQRTDNDVNIYRSIFVSGRPDDVKAHQEGDVLVWVSV
 IPEALQNDDKAKLASLTFKGNESVSTLQEQMEIQDSGMY
 AEILQINEDDKAKLASLTFKGNESVSTLQEQMEIQDSGMY
 LQSTRDYLLNNGAKAQITTDVQNDKTKVNTIDVNEGLO
 SAELEPLSLALHLNDRFRSDIADVENAIKAKLGERGYGSA
 TLVDACGRRLTVROLREFGNTVSADSLROEMROEGTWYNSQ
 ETVENRIDPINGSNDEVDVYVYKERTNGSINFGIYGTESGI
 AAVSIAGTKNDYGSVNGYETPTKDGVSGLGNPFENYDNI
 SNYTLGPVNNNSYVGLGHTYANKINFALEYNNRLYIQSMKI
 GWNYNSLNGRYFPTKGYKASLGGRVTIPGSDNKYIKLSADQ
 ADYAGFNGKRLPFYQYTVTAGGIGLRFAGYGSIGPNALYAI
 SDVIGNAIATASAEILVPPFPFVSDKSTONTVRTSLFVDAASW
 LKRLPDYCKSSRIRASTGVSFGWQSPGLPVSFAKPKKKIEN

BASE COUNT 950 a 465 c 619 g 916 t

ORIGIN

alignment_scores:
 Quality: 1052.50
 Ratio: 1.956
 Percent Similarity: 65.450
 Length: 822
 Gaps: 20
 Percent Identity: 31.630

alignment_block:
 US-09-701-711-2 x A38694 ..

Align seg 1/1 to: A38694 from: 1 to: 2950

23 SerThrHisAlaGlnAlaAspPheMetaIaAsnAspIleThrIleth 39
 ::||| ::| ||||| |||||:::|||||::: |||||:::
 373 ACAACGACTGTGTTTCCCGCACCTTTTGGCGAAAGATATTCGTGGA 422

39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPhea 56
 |||||::: |||||::: |||||::: |||||::: |||||:::
 423 TGGTGTCTCAAGGTGACTATAGAACACAAATCCGAGCAAGTTACCTGTTC 472

OM of: US-09-701-711-2 to: GenEmbl.* out_format : pfs
Date: Sep 19, 2002 4:41 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet.p2n.model -DEV=xlp
-O=/cn2.1/USPTO.spool/US09701711/runat_17092002_141431_28888/app_query.fasta_1.882
-DB=GenEmbl -QFMT=fastap -SUFFIX=sepl7oli.rge -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FCAPOP=6.000 -FCAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
-HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09701711 -CGN1_1.8909 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-701-711-2
Query length: 813
Database: GenEmbl.*
Database sequences: 1797656
Database length: 187333701
Search time (sec): 3432.230000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score.list:	Strd Orig	ZScore	Escore	Len	Documentation
gb_pat:AX009287	+ 813.00	17143.58	0.0	2442	AX009287 Sequence 1 from Patent
gb_pat:AX009289	+ 749.00	15790.34	0.0	2442	AX009289 Sequence 3 from Patent
gb_pat:AX067462	+ 749.00	15762.52	0.0	99629	AX067462 Sequence 37 from Patent
gb_pat:AX081536	+ 36.00	721.05	8.4e-32	1000	AX081536 Sequence 41 from Patent
gb_pat:AX009293	+ 11.00	213.42	0.0016	61	AX009293 Sequence 7 from Patent
gb_pat:AX009294	+ 11.00	212.72	0.0017	67	AX009294 Sequence 8 from Patent
gb_ba:PLU236920	+ 11.00	181.32	0.0976	4405	AX009294 Sequence 8 from Patent
gb_pat:AX078498	+ 10.00	166.77	0.6312	1830	AU236920 Photorehabus luminesc
gb_ba:AF120927	+ 10.00	164.63	0.8301	2433	AF120927 Shigella flexneri out
gb_ba:AF407013	+ 10.00	163.85	0.9176	2700	AF407013 Escherichia coli out
gb_ba:AE005193	+ 10.00	153.56	3.44	10649	AE005193 Escherichia coli out
gb_ba:AE004297	- 10.00	153.38	3.51	10901	AE004297 Vibrio cholerae chrc
gb_ba:AE003941	- 10.00	152.84	3.76	11711	AE003941 Xylella fastidiosa S
gb_ba:AE004784	- 10.00	152.37	4.00	12469	AE004784 Pseudomonas aerugin
gb_ba:AE000127	+ 10.00	151.05	4.74	14877	AE000127 Escherichia coli K12
gb_ba:AE008705	+ 10.00	147.78	7.20	22997	AE008705 Salmonella typhimur
gb_htg:AC104418	+ 10.00	140.33	18.73	62088	AC104418 Homo sapiens chromos
gb_htg:AC017607	+ 10.00	138.31	24.26	81255	AC017607 Drosophila melanoga
gb_ba:ECOPSF	+ 10.00	137.43	27.18	91430	D83536 Escherichia coli genom
gb_ba:ECU07214	+ 10.00	135.19	36.20	123171	U70214 Escherichia coli chrc
gb_in:AC009366	+ 10.00	132.78	49.31	169839	AC009366 Drosophila melanoga
gb_in:AC023676	+ 10.00	132.36	52.04	179611	AC023676 Drosophila melanoga
gb_ba:AJ414154	+ 10.00	131.41	58.84	204051	AJ414154 Yersinia pestis str
gb_ba:AJ414146	+ 10.00	131.19	60.50	210050	AJ414146 Yersinia pestis str
gb_ba:AL627266	+ 10.00	129.36	76.49	268050	AL627266 Salmonella enterica
gb_ba:AP002550	+ 10.00	128.99	80.19	281530	AP002550 Escherichia coli O1
gb_in:AE003597	+ 10.00	128.63	84.03	295566	AE003597 Drosophila melanoga
gb_pat:AX078597	+ 9.00	144.88	10.45	2019	AX078597 Sequence 111 from Pat
gb_pr:AY062939	+ 9.00	143.53	12.43	2418	AY062939 Homo sapiens RETSDR2
gb_pl:ZM2AG2	- 9.00	135.48	34.88	7071	X80206 Z.mays ga2 gene. 9/199
gb_pl:SPCC188	+ 9.00	123.40	164.33	35412	AL049662 S.pombe chromosome I
gb_pl:SPAC844	+ 9.00	122.84	176.49	38141	Z66569 S.pombe chromosome I
gb_pl:SPCC584	+ 9.00	122.15	192.76	41803	AL032824 S.pombe chromosome I
gb_htg:AC102899	- 9.00	118.92	291.87	64341	AC102899 Mus musculus clone F
gb_htg:AC095320	- 9.00	116.00	424.39	94948	AC095320 Rattus norvegicus cl
gb_pr:AL359454	- 9.00	115.48	453.94	101830	AL359454 Human DNA sequence
gb_htg:AC094530	+ 9.00	115.11	476.08	106997	AC094530 Rattus norvegicus c
gb_htg:AC087596	+ 9.00	114.95	485.55	109210	AC087596 Oryza sativa chromid

gb_htg:LMFLCHR16_09 - 9.00 114.90 488.93 110000 ! Continuation (10 of 11) o
gb_htg:AC016033 - 9.00 114.39 521.78 117693 ! AC016033 Homo sapiens chr
gb_pr:AL512292 + 9.00 113.34 597.05 135889 ! AL512292 Human DNA sequen
gb_pl:AP003451 - 9.00 112.57 658.99 150019 ! AP003451 Oryza sativa chr
gb_htg:AL161895 + 9.00 112.39 673.97 153566 ! AL161895 Homo sapiens chr
gb_htg:AC106473 + 9.00 112.27 684.96 156169 ! AC106473 Rattus norvegicu
gb_htg:AC018550 - 9.00 111.96 712.95 162807 ! AC018550 Homo sapiens chr

seq_name: gb_pat:AX009287

seq_documentation_block:
LOCUS AX009287 2442 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 1 from Patent WO9963093.
ACCESSION AX009287
VERSION AX009287.1 GI:9996618
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.

REFERENCE 1 (bases 1 to 2442)
AUTHORS Vinals-Bassols,C.
TITLE Basb027 proteins and genes from moraxella catarrhalis, antigens,
antibodies, and uses
JOURNAL Patent: WO 9963093-A 1 09-DEC-1999;
SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)
FEATURES
source Location/Qualifiers
i. 2442
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"

BASE COUNT 737 a 461 c 533 g 711 t
ORIGIN

alignment_scores:	Quality:	813.00	Length:	813
	Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000	
alignment_block:	US-09-701-711-2	x AX009287		
Align seg 1/1 to: AX009287	from: 1	to: 2442		
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17				
17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAla 34				
51 ATGCGTAATTCATATTTAAAGGTTTTCAGGTGAGTGCATGCAATGACATGGC 50				
34 snAspPileThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50				
101 ATGACATTATTCATCACAGGACTACAGCGAGTGCACCATTTCAAGAGCTTACAA 150				
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67				
151 AGCGTGTCGCGTTTCGTTGGTCAAGTGGTGCAGCAAAACCAAGTGGC 200				
67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84				
201 TGATGGTGTCAAGACACTTTATGCAACAGGCAATTTTTCAGATGTGCAAG 250				
84 alTyrHisGlnGluGlyArgIleleTyrGlnValThrGluArgProLeu 100				
251 TCTATCATCAAGAGGCGGTATCTATCTATCAGTAACCGAAAGGCGGTGA 300				
101 IleAlaGluIleAsnPheGluGlyAsnArgIleProLysGluGlyLe 117				
301 ATCCCTGAGATTAATTTTGGGGCAATCGCTTAATTCAAAAGAGGTTCT 350				
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134				

|||||
351 ACAAGAGGGCTAAAAAATGCTGGCTTAGCTGTGGTCAACCACCTAAAAAC 400
134 lAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnThrIleSer 150
401 AGCCACAGTACAGATGATCGAAACGAGCTTACCAATCAATATATATCA 450
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG 167
451 CRAAGGCTATTATAATACCGAATTACTGTCAACAGACGATGCTTGATGG 500
167 yAsnArgValLysLeuAspMetThrPheLaGluGlyLysProAlaArgV 184
501 TAATCGTGTAAAGCTTGATATGACCTTGTGTAAGGTAAACCTGCACGGG 550
184 aValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
551 TGGTTGATATTATATCATTTGGCAATCAGCATTTTAGCGATCGAGATTG 600
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217
601 ATTGATGCTGCTGGGATTAAGGATATAAATCAATCCACTGCTAAAGC 650
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234
651 TGACCCGTTATACTCAAGAAAGCTGTGACCACTTTAGAGAAATTTGCGTG 700
234 lAlYsTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250
701 CTAAATATCTCAATGACGGTTGTGCGTTTTCAGATTAAAGATGCTAAG 750
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHI 267
751 CTTAATATTAAATGAAGATAAAACCGCTATCTTTGTTGAGATTTCATTGCA 800
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
801 TGAAGGTGAGCAATATCGCTTTGGACAGACACAGATTTTGGGTAAATTTAA 850
284 hTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
851 CTATATACTCAAGCAGAACTTGAGGCACTGCTTAAATTCAAAGCAGAGAA 900
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
901 GGGTTTTCACAGCCATGCTTGAGCAAAACAACAATATCAGTACCAG 950
317 sPheGlyAspAspGlyTyrTyrAlaGlnIleArgProValThrArgI 334
951 ATTTGGTGACGATGGCTATTATTATGCTCAAAATCCGCTCTGTAACACGCA 1000
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
1001 TTAATCATGAAGTCGTACGGTTGATGGAAATATTATTGACCCCTGTA 1050
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPhelystThrGl 367
1051 CACCCCTGCTATGACCGCGTATTAATTTTACAGGTAACCTTTAAGACCCA 1100
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384
1101 AGATGAAGTACTCCGTCGTGAGATGCGCAACTTGAAGGTGCGTTGGCAT 1150
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
1151 CTAAATCAAAAAATCCAGCTGCTCGTCGACGCTTGATGCGGACTGGGTTT 1200
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspGI 417
1201 TTTAAACATGTTTACCGTTGATCTGCTCCAGTACCCCAACTCACCCTGATCA 1250
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434

1251 GGTGTGATGAAATTTTGTGGTTGAAGAACAACTTCAGGATCATCAACCA 1300
434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
1301 TCGCAGCAGGCTACTCTCAAGTGGTGGTGAACCTTTTCAATTTGATGTT 1350
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
1351 TCTCAAAATAACTTTTATGGTACAGTAAGCACGTCATCAATGCTTCGTTTC 1400
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
1401 TCGCTCTGAGACCCGTCGAGGTGTATAGTTGGGTATGACCAACCCATACT 1450
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
1451 TTACCGTAAATGGCGTCTCGAAAGCTTGAGTGGCTACTATCGTAAAAACC 1500
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGI 517
1501 AAGTATGATAACAAGAACAATTAGTAATTTATGACTTCTTATGCTGG 1550
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
1551 CTCATTAAGCTATGGATATCCAAATTCATGAAATCAACGCATTAAGCTTTG 1600
534 lYLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
1601 GTCGTGATCTGACAAATACCAAGCTTCATGGCGTCTGTTTATGGGCATT 1650
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
1651 AGTAATGTCAAGCAGCTGATGGCAGATGGTGCAAAATTCAGTGGATAA 1700
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIleL 584
1701 TAATGCACTTCCTGATTTTAAGCATGATTACACACCTACATATGCCATTT 1750
584 euGlyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
1751 TSGGGTGAATATTCAAGTCTAGATCGCCCTGATTTCACCAACCAAGGC 1800
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGI 617
1801 ATGAGTCATCTGTAGATTTGACGGTTGGTTTGGTGATAAAACTCATCA 1850
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysServ 634
1851 AAAAGTGGTTTATCAAGGCAATATCTATCGCCCATTTATCAAAAAATCAG 1900
634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
1901 TCTTGGTGGATACGCCAAGTTAGGCTATGGCAATAATTTTACCATTAT 1950
651 GluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSe 667
1951 GAAATTTCTATGACGGCGCTATGTTTCGTTTCGTTGCTATGATCAATC 2000
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
2001 CTCCTTTGGTCCACGCTCACAGCCATTATTGACAGCTCGTCGTTGTCAC 2050
684 lntThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
2051 AAACCACTAGGAGAGGTGTTGGTGGTAATGCTTTGGCAACTTTTCGGC 2100
701 SerGluLeuIleLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717
2101 AGTGAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATGATCAGGT 2150
717 lArgProValIlePheIleGluGlyGlnValPheAspThrThrGlyM 734
2151 GCGTCCAGTGATATTTCATTGAGGCGGTCAGGTTTTTTGATACACAGGTA 2200

```

734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
|||||
2201 TGGATAACAACACCATGATTAAACCAATTTAAAGACCCACACACACA 2250

751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLys 767
|||||
2251 GCTGAACAAAATGCAAAAGACGCCAATCGCCGCTACTAACCCAAAGATAA 2300

767 scLnuLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784
|||||
2301 ACAGTTGGCTTATAGTGTGGTGGTGGCACTTGGTATAGCCCAATG 2350

784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
|||||
2351 GTCCCTTATCTATTAGCTATGCCAAGCCATTGAATAAAAAACAATAATGAT 2400

801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
2401 CAGACCGATACGGTACAGTTCAGATTGGTAGTGTCTTT 2439

seq_name: gb_pat:AX009289
seq_documentation_block:
LOCUS AX009289 2442 bp DNA linear PAT 06-SEP-2000
DEFINITION Sequence 3 from Patent WO9963093.
ACCESSION AX009289
VERSION AX009289.1 GI:9996619
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE 1 (bases 1 to 2442)
AUTHORS Vinals-Bassols,C.
TITLE Bas027 proteins and genes from moraxella catarrhalis, antigens,
antibodies, and uses
JOURNAL Patent: WO 9963093-A 3 09-DEC-1999;
SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)
FEATURES
Location/Qualifiers
source 1..2442
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
BASE COUNT 735 a 461 c 535 g 711 t
ORIGIN

alignment_scores:
Quality: 749.00 Length: 749
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AX009289 ..
Align seg 1/1 to: AX009289 from: 1 to: 2442

65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAs 81
|||||
193 CAGTTGGCTGATGGTGTCAAGACACTTTATGCAACAGGCAATTTTCAGA 242

81 pValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluA 98
|||||
243 TGTGCAAGTCTATCATCAAGAGGCGGTATCATCTATCAGGTAAACCGAAA 292

98 rgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLys 114
|||||
293 GGCCGTTAATCGCTGAGATTAATTTGAGGGCAATCGCTTAATTCAAAAA 342

115 GluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131
|||||
343 GAAGGTCTACAAGAGGGGCTAAAAAATGCTGGCTAGTGTGGGTCAAC 392

```

```

131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148
|||||
393 ACTAAACAAGCCACAGTACAGATGATCGAAACCGAGCTTACCATCAAT 442

148 yrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMet 164
|||||
443 ATATATCAACAAGGCTATTATATACCAAAATTTACTGTCAACACAGCATG 492

165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181
|||||
493 CTGTGATGTAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAC 542

181 oAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspA 198
|||||
543 TGCACGGGTGGTGTGATATTATATCATTTGGCAATCAGCATTTTAGCGATG 592

198 laAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsnProLeu 214
|||||
593 CAGATTTGATTGATGCTGCTGCGATTAAAGGATAATAAAATCAATCCACTG 642

215 SerLysAlaAspArgTyrThrGlnLysLeuValThrSerLeuGluAs 231
|||||
643 TCTAAAGCTGACCGTTATCTCAACAAAGCTGGTGACCACTTTAGAGAA 592

231 nLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysA 248
|||||
693 TTTGCGTGTCTAAATATCTCAATGCAGGGTTTGTGCGTTTGGATTAAG 742

248 spAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIle 264
|||||
743 ATGCTAAAGCTTAATATTAAATGAAGATAAAACCGTATCTTTGTTGAGAT 792

265 SerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGl 281
|||||
793 TCATTGCAATGAAGGTGAGCAATATCGCTTGGACAGACACACTTTTGGG 842

281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysA 298
|||||
843 TAATTTAACTTATCTCAAGCAGAACTTTGAGGCACTGCTTAAATTCAAA 892

298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314
|||||
893 CAGAAGAAGGGTTTTCACAAGCCATGCTTGAGCAACAACAACAATATC 942

315 SerThrLysPheGlyAspGlyTyrTyrAlaGlnIleArgProVa 331
|||||
943 AGTACCAAAATTTGGTGACGATGGCTATTATATGCTCAAAATCCGCTCTGT 992

331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348
|||||
993 AACCGCATTAATGATGAAGTCGTACGGTTGATGTGGAATATTATATTG 1042

348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364
|||||
1043 ACCCTGTACACCCCTGCTATGTACGCCGTATTAAATTTTACAGSTAAC 1092

365 LysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381
|||||
1093 AAGACCCCAAGATGAAGTACTCGTCTGAGATGCCAACAATCTGAAGGTGC 1142

381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398
|||||
1143 GTTGGCATCTAATCAAAAAATCCACGCTGCTCGTCACGCTTGATGCGGA 1192

398 hrGlyPhePheLysHisValThrValAspThrArgProValProAsnSer 414
|||||
1193 CTGGGTTTTTAAACATGTTACCGTTGATCTCGTCCAGTACCCCACTCA 1242

415 ProAspGlnValAspValAsnPheValValGluGlnProSerGlySe 431
|||||
1243 CCTGATCAGGTTGATGTAATTTTGTGGTTGAAGAACCAACCTTCAGGATC 1292

431 rSerThrIleAlaAlaGlyTyrSerGlnSerGlyValThrPheGlnP 448

```

```
|||||
1293 ATCAACCATCGCAGCAGGTACTCTCAAGAGTGGTGTAACCTTTTCAAT 1342
|||||
448 heAspValSerGlnAsnAsnPhMetGlyThrGlyLysHisValAsnAla 464
|||||
1343 TTGATGTTTCTCAAAATAACTTTATGGGTACAGTAAGCAGCGTCAATGCT 1392
|||||
465 SerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAs 481
|||||
1393 TCGTTTCTCGCTCTGAGACCCGTGAGGTGTATAGTTTGGGTATGACCAA 1442
|||||
481 nProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrA 498
|||||
1443 CCCATACTTTACCGTAAATGGCGCTCTCGCAAGCCTTGAGTGGCTACTATC 1492
|||||
498 rGlyThrLysTyrAspAsnLysAsnLysSerAsnTyrValLeuAspSer 514
|||||
1493 GTAAACCAAGTAGTATGATAACAAGAACATAGTAAATTATGCTATGATCT 1542
|||||
515 TyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIl 531
|||||
1543 TATGGTGGCTCATAGCTATGATATCCAAATGATGAATAATCAACGCAT 1592
|||||
531 eSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheM 548
|||||
1593 AAGCTTTGGTCTCAATGCTGACAATACCAAGCTTCATGGCGGTGCTTTTA 1642
|||||
548 etGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGln 564
|||||
1643 TGGGCATTAGTAATGTCACAGCCTCATGGCAGATGGTGCAAAATTCAA 1692
|||||
565 ValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrTyrAs 581
|||||
1693 GTGGATAAATAATGCAATCTCTGATTTTAAGCATGATTACACAACCTACAA 1742
|||||
581 nAlaIleLeuGlyTyrPAsnTyrSerSerLeuAspArgProValPheProT 598
|||||
1743 TGGCATTTTGGGTGGAAATATTATCAAGTCTAGATCGCCCTGTATTCCAA 1792
|||||
598 hrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAspLys 614
|||||
1793 CCCAAGCATGAGTCATCTGTAAGATTTGACGGTTGGTTTGGTGATATAA 1842
|||||
615 ThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLy 631
|||||
1843 ACTCATCAAAAGTGGTTTATCAAGCAATATCTATCGCCCATTTATCAA 1892
|||||
631 sLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeup 648
|||||
1893 AAATCAGTCTTGGGTGATACCCCAAGTTAGGCTATGGCAATAATTTAC 1942
|||||
648 roPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyr 664
|||||
1943 CATTTATGAAATTTCTATGACGGCGGTATGGTTCGGTTCGTGGCTAT 1992
|||||
665 AspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgAr 681
|||||
1993 GATCAATCCCTCTTTGGGTCCACGCTCACAAAGCCTATTGACAGCTCGTCG 2042
|||||
681 qGlyGlnGlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaT 698
|||||
2043 TGGTCAACAAACCACTAGGAGAGGTGTGGTGGTAAATGCTTTGGCAA 2092
|||||
698 hrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTrpIle 714
|||||
2093 CTTTCGGCAGTAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATA 2142
|||||
715 AspGlnValArgProValIlePheIleGluGlyGlnValPheAspTh 731
|||||
2143 GATCAGGTGGCTCAGTGATATTCAVTTAGGCGGTCAGGTTTTTTGATAC 2192
|||||
731 rThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProG 748
|||||
```

```
2193 AACAGGTATGATAAACAACCACTTGATTTAACCACTTTAAAGACCCAC 2242
|||||
748 InAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThr 764
|||||
2243 AAGCAACAGCTGAACAAATGCAAAAGCAGCCCAATCGCCGCTACTAACCC 2292
|||||
765 GlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrTh 781
|||||
2293 CAAGATAAACAGTTGCGTTATAGTGGTGGTGGTCAACTTGGTATAC 2342
|||||
781 rProLleGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysG 798
|||||
2343 GCCCATTTGGTCTTTATCTATTAGCTATGCAAGCCATTGAATAAAAAAC 2392
|||||
798 InAsnAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
2393 AAAATGATCAGACCGATACGGTACAGTTCCAGATTGGTAGTGTCTTT 2439
|||||
seq_name: gb_pat:AX067462
seq_documentation_block:
LOCUS AX067462 99629 bp DNA linear PAT 24-JAN-2001
DEFINITION Sequence 37 from Patent WO0078968.
ACCESSION AX067462
VERSION AX067462.1 GI:12545082
KEYWORDS
SOURCE Moraxella catarrhalis.
ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.
REFERENCE
1 (bases 1 to 99629)
AUTHORS Lagace,R.E., Patterson,C. and Berg,K.L.
TITLE Nucleotide sequences of moraxella catarrhalis genome
JOURNAL Patent: WO 0078968-A 37 28-DEC-2000;
Incyte Genomics, Inc. (US)
FEATURES
location/Qualifiers
source 1..99629
/organism="Moraxella catarrhalis"
/db_xref="taxon:480"
78467
/note="a or g or c or t, unknown, or other"
BASE COUNT 29233 a 19222 c 21909 g 29264 t 1 others
ORIGIN
alignment_scores:
Quality: 749.00 Length: 749
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x AX067462 ..
Align seg 1/1 to: AX067462 from: 1 to: 99629
65 GlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAs 81
|||||
66325 CAGTTGGCTGATGGTGCAAGACACTTTATGCAACAGGCAATTTTCAGA 66374
|||||
81 pValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluA 98
|||||
66375 TGTGCAAGTCTATCATCAAGAGGGCGTATCATCTATCAGGTACCGGAA 66424
|||||
98 rgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLys 114
|||||
66425 GGCCGTAAATCGCTGAGATTAATTTGAGGCAATCGCTTAATTCACAAA 66474
|||||
115 GluGlyLeuGlnGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnPr 131
|||||
66475 GAAGTCTACAAGAAGGGCTAAAAAATGCTGGCTTAGCTTGGGTCAACC 66524
|||||
131 oLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnT 148
|||||
```

```

66525 ACTAAACAAGCCACGACGATGATCGAAACCGAGCTTACCAATCAAT 66574
148 yrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMet 164
66575 ATATATCACAGGCTATTATTAATACCGAATTAAGTCTCAACACACAGG 66624
165 LeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysPr 181
66625 CTTGATGTAATCGTGTAAAGCTTGATGATGACCTTGTCTGAAGGTAAAC 66674
181 oAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspA 198
66675 TGCACGGGTGGTGTATTAATATCATTTGGCAATCAGCATTTTACGATG 66724
198 laAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsnProLeu 214
66725 CAGATTTGATTTGATGCTGCTGCGATTAAAGGATAATAAATCAATCCAC 66774
215 SerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAs 231
66775 TCTAAAGCTGACCTTATCTCAAGAAAGCTGGTGACCACTTTAGAGAA 66824
231 nLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLys 248
66825 TTTGCGTGCTAAATATCTCAATGACGGGTTTGTGCGTTTGAGATTAAAG 66874
248 spAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIle 264
66875 ATGCTAAGCTTAATTAATTAAGATTAATAAACCGTATCTTTGTTGAGAT 66924
265 SerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGl 281
66925 TCATTTGATGAAGTGAGCAATATCGCTTTGGACAGACACAGTTTGTGGG 66974
281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLys 298
66975 TAATTTAACTATATCTCAAGCAGAACTTGAGGCACCTGCTTAAATTCAAAG 67024
298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314
67025 CAGAAGAAGGTTTTCACAGCCATGCTTGAGCAACAACAACAAATATC 67074
315 SerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProVa 331
67075 AGTACCAAAATTTGTCGACGATGCTATTATATGCTCAAAATCCGCTCGT 67124
331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348
67125 AACACGCAATTAATGATGAAAGTCGTACGGTTGATGTGGAATATATATG 67174
348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364
67175 ACCCTGTACACCCCTGTCTATGTACGCCGTATTAATTTTACAGGTAAC 67224
365 yLysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381
67225 AAGACCAAGATGAAGTACTCCGTCGTGAGATGCGACAACTTGAAGTGC 67274
381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398
67275 GTTGGCATCTAAATCAAAAAATCCAGCTGTCTCGTGCACGCTTGATGCG 67324
398 hrGlyPhePheLysHisValThrValAspThrArgProValProAsnSer 414
67325 CTGGGTTTTTAAACATGTTTACCGTTGATACTCGTCCAGTACCCCACTCA 67374
415 ProAspGlnValAspValAsnPheValValIcIuGluGlnProSerGlySe 431
67375 CCTGATCAGGTTGATGTAATTTTGTGGTTCAGAAACAACCTTCCAGGATC 67424
431 rSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnP 448
67425 ATCAACCATCGCAGCAGGCTACTCTCAAGAGTGGTGGTGAACCTTTTCAAT 67474

```

```

448 heAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAla 464
67475 TTGATGTTTCTCAAAATAACTTTATGGGTACAGGTAAACGACGTCAATGCT 67524
465 SerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAs 481
67525 TCGTTTTCTCGCTCTGAGACCCGTGAGGTATAGTTTGGGTATGACCAA 67574
481 nProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrA 498
67575 CCCATACCTTTACCCGTAAATGGCGTCTCGCAAGCTTGAGTGGCTACTATC 67624
498 rGlyLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSer 514
67625 GTAAACCAAGTATGATNAACAAGAACAATTAGTAATTATGTACTTGATTCT 67674
515 TyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIl 531
67675 TATGGTGGCTCATTAAGCTATGGATATCCAATTGATGAAAAATCAACGCAT 67724
531 eSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheM 548
67725 AAGCTTTGGTCTGAATGCTGACAATACCAAGCTTCATGGCGGTCTGTTTA 67774
548 etGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGln 564
67775 TGGCATATTAGTAATGTCAAGCAGCTGATGGCAGATGGTGGCAAAATTCAA 67824
565 ValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrTyrAs 581
67825 GTGGATAATAATGTCATTTCCTGATTTTAAAGCATGATTACACAACCTACAA 67874
581 nAlaIleLeuGlyTyrPasnTyrSerSerLeuAspArgProValPheProT 598
67875 TGCCATTTGGGGTGAATTAATCAAGTCTAGATCGCCCTGTATTTCCAA 67924
598 hrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAspLys 614
67925 CCCAAGCAGTACGTCATCTGTAGATTTGACGGTTGGTTTTGGTGATAAA 67974
615 ThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLy 631
67975 ACTCATCAAAAAGTGGTTTATCAAGGCAATATCTATCGCCCATTTATCAA 68024
631 sLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuP 648
68025 AAAATCAGTCTTGGCTGGATACGCCAAGTTAGGCTATGGCAATAATTTAC 68074
648 roPheTyrGluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyr 664
68075 CATTTTATGAAAAATTTCTATCGACGGCGCTATGGTTCGGTTCGGGCTAT 68124
665 AspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgAr 681
68125 GATCAATCCTCTTTGGGTCCACGCTCAACAGCCTATTTGACAGCTCGCTG 68174
681 gGlyGlnGlnThrThrLeuGlyGluValValIcIyGlyAsnAlaLeuAlaT 698
68175 TGGTCAACAAACCACTAGGAGAGGTTGTTGGTGGTAATGCTTTGGCAA 68224
698 hrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTrrIle 714
68225 CTTTCGGCAGTACGCTGATTTTACTTTTGCCTTTAAAGGTGATTTGGATA 68274
715 AspGlnValArgProValIlePheIleGluGlyGlnValPheAspTh 731
68275 GATCAGGTGCGTCCAGTATATTCATTGAGGGCGGTCAAGTTTTTGTATAC 68324
731 rThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLysAspProG 748
68325 AACAGGTATGGATAACAACAACCATTTGATTTAACCCCAATTTAAAGACCCAC 68374

```

748 lnaAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThr 764
|||||
68375 AAGCAACAGCTGAACAAATGCAAAAGCAGCCAAATCGCCGCTACTAAC 68424
|||||
765 GlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrTh 781
|||||
68425 CAAGATAAACAGTGGCTTATAGTGTGGTGGTGGCAACTGGGTATAC 68474
|||||
781 rProLeGlyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysG 798
|||||
68475 GCCCATGTGCTTATCTATATAGCTATGCCAAGCCATTTGAATAAAAC 68524
|||||
798 lnaAsnAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
68525 AAAATGATCAGACCGATACGATCGGTACAGTTCAGATTGGTAGTGCTTT 68571
|||||

seq_name: gb_pat:AX081536

seq_documentation_block:
LOCUS AX081536 1000 bp DNA linear PAT 27-FEB-2001
DEFINITION Sequence 41 from Patent WO0109350.

ACCESSION AX081536
VERSION AX081536.1 GI:13170355

KEYWORDS
SOURCE Moraxella catarrhalis.

ORGANISM Moraxella catarrhalis
Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
Moraxella.

REFERENCE 1 (bases 1 to 1000)

AUTHORS Berthet,F.X., Dalemans,W.L., Denoel,P., Dequesne,G.S., Peron,C.S.,
Lobet,Y.S., Poolman,J.S., Thiry,G.S., Lhonnard,J.S. and Voet,P.S.,
Genetically engineered bleb vaccine

TITLE Patent: WO 0109350-A 41 08-FEB-2001;

JOURNAL SMITHKLINE BEECHAM BIOLOGICALS S.A. (BE)

FEATURES
Location/Qualifiers

source
1..1000
/organism="Moraxella catarrhalis"

/db_xref="taxon:480"

BASE COUNT 287 a 172 c 215 g 326 t

ORIGIN

alignment_scores:
Quality: 36.00 Length: 36
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AX081536 ..

Align seg 1/1 to: AX081536 from: 1 to: 1000

1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetal 17
|||||

821 ATGGGTAATTCATATTTAAAGGTTTCAGGTGAGTGCATGACATGGC 870
|||||

17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetalA 34
|||||

871 TGTCATGATGTAATGCACTCATGCACAGCGCGGATTTTATGGCAA 920
|||||

34 snAspIle 36
|||||

921 ATGACATT 928

seq_name: gb_pat:AX009293

seq_documentation_block:
LOCUS AX009293 61 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 7 from Patent WO9963093.

ACCESSION AX009293

VERSION AX009293.1 GI:9996622

KEYWORDS
SOURCE synthetic construct.

ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 61)
AUTHORS Vinals-Bassols,C.
TITLE Basb027 proteins and genes from moraxella catarrhalis, antigens,
antibodies, and uses
JOURNAL Patent: WO 9963093-A 7 09-DEC-1999;
SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)
FEATURES
Location/Qualifiers
source
1..61
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="Oligonucleotide"

BASE COUNT 20 a 15 c 17 g 9 t
ORIGIN

alignment_scores:
Quality: 11.00 Length: 11
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AX009293 ..

Align seg 1/1 to: AX009293 from: 1 to: 61

39 ThrGlyLeuGlnArgValThrIleGlySerLeu 49
|||||

28 ACAGGACTACAGCGAGTGACCATTTGAAAGCTTA 60
|||||

seq_name: gb_pat:AX009294

seq_documentation_block:
LOCUS AX009294 67 bp DNA linear PAT 06-SEP-2000

DEFINITION Sequence 8 from Patent WO9963093.

ACCESSION AX009294

VERSION AX009294.1 GI:9996623

KEYWORDS
SOURCE synthetic construct.

ORGANISM synthetic construct.
artificial sequence.

REFERENCE 1 (bases 1 to 67)

AUTHORS Vinals-Bassols,C.

TITLE Basb027 proteins and genes from moraxella catarrhalis, antigens,
antibodies, and uses
JOURNAL Patent: WO 9963093-A 8 09-DEC-1999;
SMITHKLINE BEECHAM BIOLOG (BE); VINALS BASSOLS CARLOTA (BE)

FEATURES
Location/Qualifiers
source
1..67
/organism="synthetic construct"

/db_xref="taxon:32630"

/note="Oligonucleotide"

BASE COUNT 22 a 16 c 15 g 14 t
ORIGIN

alignment_scores:
Quality: 11.00 Length: 11
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AX009294/rev ..

Align seg 1/1 to reverse of: AX009294 from: 1 to: 67

803 AspThrValGlnPheGlnIleGlySerValPhe 813
|||||

66 GATACGGTACAGTTCCAGATTGGTAGTGCTTT 34
|||||

seq_name: gb_ba:PLU236920

Nelson, K.E., Hickey, E. and Peterson, J.
 Dna sequences which are suited for specifically detecting
 pseudomonas putida kt2440
 Patent: WO 0107624-A 12 01-FEB-2001;
 THE INSTITUTE FOR GENOMIC RESEARCH (US) ; OIAGEN GmbH (DE) ;
 Gesellschaft fuer Biotechnologische Forschung mbH (GBF) (DE) ;
 Deutsches Krebsforschungszentrum (DKFZ) (DE) ; Medizinische
 Hochschule Hannover (DE)
 FEATURES
 source Location/Qualifiers
 1..1830
 /organism="Pseudomonas putida"
 /db_xref="taxon:303"
 BASE COUNT 401 a 555 c 529 g 344 t 1 others
 ORIGIN

alignment_scores:
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-701-711-2 x AX078498 ..
 Align seg 1/1 to: AX078498 from: 1 to: 1830
 368 AspGluValLeuArgGluMetArgGln 377
 |||||
 1081 GAGCAAGTGTGGTGGCGAATGCCCGCAG 1110

seq_name: gb_ba:AF120927

seq_documentation_block:
 LOCUS AF120927 2433 bp DNA linear BCT 02-MAR-2001
 DEFINITION Shigella flexneri outer membrane antigen Oma90 (oma90) gene,
 complete cds.
 ACCESSION AF120927
 VERSION AF120927.1 GI:4567045
 KEYWORDS
 SOURCE Shigella flexneri.
 ORGANISM Shigella flexneri
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Shigella.
 REFERENCE 1 (bases 1 to 2433)
 Robb, C.W., Orihuela, C.J., Ekkelenkamp, M.B. and Niesel, D.W.
 Identification and characterization of an in vivo regulated
 D15/Oma87 homologue in Shigella flexneri using differential display
 polymerase chain reaction
 JOURNAL Gene 262 (1-2), 169-177 (2001)
 PUBMED 11179681
 REFERENCE 2 (bases 1 to 2433)
 Robb, C.W., Ekkelenkamp, M.B., Orihuela, C.J. and Niesel, D.W.
 Direct Submission
 TITLE Submitted (16-JAN-1999) Microbiology and Immunology, University of
 Texas Medical Branch, 301 Administration Blvd, Galveston, TX 77555,
 USA
 FEATURES
 source Location/Qualifiers
 1..2433
 /organism="Shigella flexneri"
 /db_xref="taxon:623"
 1..2433
 /gene="oma90"
 1..2433
 /gene="oma90"
 /note="similar to Escherichia coli hypothetical YaeT, 90
 kDa"
 /codon_start=1
 /transl_table=11
 /product="outer membrane antigen Oma90"
 /protein_id="AAD23568.1"
 /db_xref="GI:4567046"
 /translation="MAMKLLIASLFFSSATVYAGEGFVVKDIIHFGLQRVAVGAALL
 SMPVRTGDTVNDISNTIRALFATGNFEDVRLRDGDTLLVQVKERTIASITFSGN

KSVKDDMLKQNLASGYRVGESLDRTTIADIEKGLEDFYISVGKYSASVKAIVTLP
 NRVDKLVEGVSAAEQIINIGNHAFITDDLSHFQIVNITEGDTKLKSGVEVSGK
 LAGDLETLRYSYLDGRYAFNRINDTSQVSLTPDKKGIYVTVNITGDTKLKSGVEVSGK
 LAGHSAIEQLTKIEPGELYNGTKVTRKEDDIKKLGRYGYAVPRVQSMPEINDAKT
 VKLRVNDAGNRVYKIRFEGNTSKDAVLRREMRQMEGAWLGSDLVDQGERLNRLL
 GFPEVDTQRPVSPDQVDVYKVERNTGSFNGFYGCTGSEGSFQAGVQODNWLL
 GTGVAVGINTGKNDYQYIAELSVNPFYTVDCVSLGRLFYNDQFQADADLSDYTNKS
 YGTDVTIGFPINEYNSRAGLGYVHNSLNMQPVAMWRYLISMEHPSFTSDQDNQFK
 TDDFTNYGWTYINKLDRGYPTDGSRYNLTGKVTIPGSDNEYIYKTLDTATYVPIDDDD
 HKWVLTGRWYGDGLGKEMPFYFAGGSSTVRGFSQNTIGPKAVYFPHQASNY
 DPDYDECATODGAKDLCKSDDAVGNAMAVASLEFIPTPTFIDSKVANSVRTSFFWD
 MGTWDTWDSOYSGYPDYSNIRMSAGIALQWMSPLGLVFSYAQPKKYDGDK
 AEQFPNIGKTW"
 BASE COUNT 606 a 586 c 647 g 594 t
 ORIGIN

alignment_scores:
 Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-701-711-2 x AF120927 ..
 Align seg 1/1 to: AF120927 from: 1 to: 2433
 648 ProPheTyrGluAsnPheTyrAlaGlyGly 657
 |||||
 1939 CCGTCTCAGAGAACTTCTATGCCGGTGT 1968

seq_name: gb_ba:AF407013

seq_documentation_block:
 LOCUS AF407013 2700 bp DNA linear BCT 10-SEP-2001
 DEFINITION Escherichia coli outer membrane protein (ecfK) gene, complete cds.
 ACCESSION AF407013
 VERSION AF407013.1 GI:15529631
 KEYWORDS
 SOURCE Escherichia coli.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 2700)
 Dartigalongue, C., Missiakas, D. and Raina, S.
 Characterization of the Escherichia coli sigma E regulon
 J. Biol. Chem. 276 (24), 20866-20875 (2001)
 MEDLINE 21293005
 PUBMED 11274153
 REFERENCE 2 (bases 1 to 2700)
 Dartigalongue, C., Missiakas, D. and Raina, S.
 Direct Submission
 TITLE Submitted (09-AUG-2001) Biochimie Medicale, CMU, 1 Rue Michel
 Servet, Geneva 1211, Switzerland
 JOURNAL Location/Qualifiers
 FEATURES
 source
 1..2700
 /organism="Escherichia coli"
 /db_xref="taxon:562"
 61..>2640
 /gene="ecfK"
 61..100
 /gene="ecfK"
 /note="factor sigma24"
 101..>2640
 /gene="ecfK"
 /note="outer membrane protein"
 208..2640
 /gene="ecfK"
 /note="Ecfc; under heat shock transcriptional control of
 RpoE; not essential for Escherichia coli growth; similar
 to Oma87 and D-15 antigen"
 /codon_start=1
 /transl_table=11

```
/product="outer membrane protein"
/protein_id="AAL01379.1"
/db_xref="GI:1529632"
/translation="MAWKLLIASLLFSSATVYGAEGFVVVKDITHEGLQRAVGAALL
SMPTGCTVDVNDISNTIRALFATGNFEDRVLRDGTLLVQVKRPTIASITFSGN
KSVKDKMLKONLEASGVVRGESLDRTTIADIEKGLDFYISVGKYSASVKAATVPLPR
NRVDLKLFOEGYSAEIQIINIVGNHAFTTDELISHFOLRDEVPWNVVGDRYQOK
LAGDLTSLSYLDRGYARENDISTOVSLTPDKKGIYVTVNITEGQYKLSGVEVSGN
LAGSAREIOLTKIEGELYNGTKYKMEDDIKLLGRGYAYPRVQSPEINDADKT
VKLURNVAGNRYIVRKIRFEGNDTSKDAVLRMRQMEGAWLGSDLVDGKERLNLK
GFETVDTQTRVPGSPQDVVYKVKERTGNTSFNGIGYGTESGVFQAGVQODNML
GTGYAVGINKNDQTYAELSVNTNFTYVDGVSGLGRFLYNDFOADDLSDTNKS
YGVDTVLGFPINENSLRAGLVHNSLMQPOVAMRWLYLSMGHPSTSDQDNFSK
TDDFTNGYKTYNKLDRGYEPTDGSRVNLTKVTIPCSONEYKVTLDATATYVPIDD
HKWYLGTRPMGVDGLGKEMPEFYENYAGGSSTVRGFSQNTIGPKANYFPHQASNY
DPDIECATQDCAKDLCKSDDAVAGGNAMVASLEFIPTPFIISDKTANSVTSFWD
MGVTWDTNDSSQSYGPDYSDPSNIRMSAGIALQWNSPLGLVFSYAOPFKKYDGGK
AEQFQFNIGKTW"
BASE COUNT      668 a   636 c   721 g   675 t
ORIGIN

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-701-711-2 x AF407013 ..

Align seg 1/1 to: AF407013 from: 1 to: 2700

648  ProPheTyRGluAspPheTyRAlaGlyGly 657
      |||||
2146  CGGTCTACGAGAACTATGCGCGTGT 2175

seq_name: gb_ba:AE005193

seq_documentation_block:
LOCUS      AE005193              10649 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 17
of 155.
ACCESSION      AE005193 AE005174
VERSION        AE005193.1 GI:12512893
KEYWORDS
SOURCE        Escherichia coli O157:H7 EDL933.
  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Escherichia.
REFERENCE      1 (bases 1 to 10649)
  Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
  Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
  Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
  Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
  Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
  Welch,R.A. and Blattner,F.R.
  Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  Nature 409 (5819), 529-533 (2001)
  21074935
  Nature 409 (5819), 529-533 (2001)
  11206551
REFERENCE      2 (bases 1 to 10649)
  Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
  Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
  Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
  Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
  Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
  Welch,R.A. and Blattner,F.R.
  Direct Submission
  Submitted (22-OCT-2000) Laboratory of Genetics, University of
  Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
  Location/Qualifiers
    1..10649
    /organism="Escherichia coli O157:H7 EDL933"

/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
215..940
/gene="rpsB"
/note="20180"
215..940
/gene="rpsB"
/function="structural component; Macromolecule synthesis,
modification: Ribosomal proteins - synthesis,
modification:riboso"
/note="Residues 1 to 241 of 241 are 100.00 pct identical
to residues 1 to 241 of 241 from Escherichia coli K-12
Strain MG1655: B0169"
/codon_start=1
/transl_table=11
/product="30S ribosomal subunit protein S2"
/protein_id="AAG54471.1"
/db_xref="GI:12512894"
/translation="MATVSMRDLKAGVHFGHOTOYWNPKMPFIFGARNKVHIINLE
KTYPMNEALAEALNKIASRKGLFVCTKRAASAVDAALSCDQFFVNHFWLGMILT
NWKTVROSIRLKDLETQSDGTFDKLTKEALMRTRELEKLENSLGKIDKMGSLPDA
LFVIDADHEHIAIKEANLGIPIVFAIVDTNSDPDGVDFVIPGNDADRAVTLYLGA
ATVREGRSODLASQAEESEVEAE"
1198..2049
/gene="tsf"
/note="20181"
1198..2049
/gene="tsf"
/function="factor; Macromolecule synthesis, modification:
proteins - translation and modification"
/note="Residues 1 to 283 of 283 are 100.00 pct identical
to residues 1 to 283 of 283 from Escherichia coli K-12
Strain MG1655: B0170"
/codon_start=1
/transl_table=11
/product="protein chain elongation factor EF-Ts"
/protein_id="AAG54472.1"
/db_xref="GI:12512895"
/translation="MAEITASLVKELRERTGAGMDCCKKALTEANGDIELAIENWRKS
GATKAAGKAGNVAADGVIKTKIDNGYGIILEVNCQTFVAKDGFQAFADKVLDAAG
KGITDVELKAEFEERVALVAKIGENIRRAALEGDLVGSYQHGARGVLVAAG
ADELVKHIAMHVAASKPEFTKPEDVSAEVVEYQVOLDIAMQSGPKKEIAERWEG
RMKKTGEVSLTGQPFVMEPSKTVGLLKEHNAEVTGFIREFEGEIKVETDFAAEV
AAMSKQS"
2196..2921
/gene="pyrH"
/note="20182"
2196..2921
/gene="pyrH"
/function="enzyme; Central intermediary metabolism:
Nucleotide interconversions"
/note="Residues 1 to 241 of 241 are 100.00 pct identical
to residues 1 to 241 of 241 from Escherichia coli K-12
Strain MG1655: B0171"
/codon_start=1
/transl_table=11
/product="uridylylate kinase"
/protein_id="AAG54473.1"
/db_xref="GI:12512896"
/translation="MATNAKPVYKRILLKLSGALQGTGEGIDASILDRMAQEIKEK
VELGIQGVVIGGNLFRGAGLAKAGNRVVGDMGLMATVNGDLAMRDALHRAVVA
RLMSAIPLNGVCSYSWAEALSLRNRRVTLISAGTGNPFTTDSACLRCGIEADY
VLKATKVDGVFTADPAKDPATMYEQITYSEVLEKELKVMDLAAFTLARDHKLPIRV
NMKNKPGALRRVYMGKEGEGTLITE"
3071..3628
/gene="frr"
/note="20183"
3071..3628
/gene="frr"
/function="factor; Macromolecule synthesis, modification:
Proteins - translation and modification"
```

/note="Residues 1 to 185 of 185 are 100.00 pct identical to residues 1 to 185 of 185 from Escherichia coli K-12 Strain MG1655: B0172"
Strain MG1655: B0172"
/codon_start=1
/transl_table=1
/product="ribosome releasing factor"
/protein_id="AAG54474.1"
/db_xref="GI:12512897"
/translation="MISDIRKDAEVRMDKCVFAFKTQISKIRTSRASPSSLLDGIIVVEY YGTPPLQLASVTVEDSRTLKINVFDRSMSPAVEKAIMASDLGLNPNAGSDIRVPL PPLTERKDKLTKIVRGEAEOARVAVRNRDANDKVKALLDKKEISDDDDRRSDDVV OKLTDAAIKKIEAALADKAEALMQF"
3720..4916
/gene="yaem"
/note="20184"
3720..4916
/gene="yaem"
/function="putative transport; Not classified"
/note="Residues 1 to 398 of 398 are 99.74 pct identical to residues 1 to 398 of 398 from Escherichia coli K-12 Strain MG1655: B0173"
MG1655: B0173"
/codon_start=1
/transl_table=1
/product="putative ATP-binding component of a transport system"
/protein_id="AAG54475.1"
/db_xref="GI:12512898"
/translation="MKQLTILGSTGICGSTDVVVRHNPFRVVAALVAGKNVTRMVE OCLEFSPYAVMDDEASAKLLTLMQOGSRTFVLSGGOACDMALEDVQVMAIV GAAGLLPTLAIRAGKTILLANKESLVTGRLFMDAVKQSKAQLLPVSEHNAIPQSL PQPIOHNLGYADLEONGVSVILLTSGSGGPFRETPLRDLATMPDQACRHPNWSMGRKI SVDSATMKNKGLXEYLEARWLFNASQOMEVLHPQSVIHSVMRYQDGSVLAQLGEPDM RTPAIHTAWPNRVNSGVKPLDFCKLSALTAAPDYDRYPCCLKLAMEAFEGQQAATTA LNAANEITVAFAFLAQOIRFTDIAALNLSVLEKMDMREPQCVDVLSVDASAREVAKK VMLAS"
5102..5863
/gene="yaes"
/note="20185"
5102..5863
/gene="yaes"
/function="orf; Unknown function"
/note="Residues 1 to 253 of 253 are 100.00 pct identical to residues 1 to 253 of 253 from Escherichia coli K-12 Strain MG1655: B0174"
Strain MG1655: B0174"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG54476.1"
/db_xref="GI:12512899"
/translation="MMLSATQPLSEKLPAGCRHVAIIMDNGRWAKQKIRAFGHK AGAKSVRAVSFAANNQIEALTLYAFSENWNRPAQVYSALMELFVWALDSEVKSLSHR HNVRLRIIGDTSRFNSRLQERIRKSEALTAGNTGLTINIAANYGRWMDIVQGVQLAE KVQCNLQPDQIDEMLNOHVMHAPDLVIRTGGEHRI SNELLWQIAVYAELYFTD VLPDPDFQDFEGALNAFANRRERFGGTEPGDETA"
5984..6733
/gene="cdsa"
/note="20186"
5984..6733
/gene="cdsa"
/function="enzyme; Fatty acid biosynthesis: Fatty acid and phosphatidic acid biosynth"
/note="Residues 1 to 249 of 249 are 100.00 pct identical to residues 1 to 249 of 249 from Escherichia coli K-12 Strain MG1655: B0175"
Strain MG1655: B0175"
/codon_start=1
/transl_table=1
/product="CDP-diglyceride synthetase"
/protein_id="AAG54477.1"
/db_xref="GI:12512900"
/translation="MLAAEWGQLSGFTTSRQVRVLAIVLCGLLLALMLFLPEYHRNI HOPVLVSLWASGLWIVALLLVLFPGSAAIWRNSKTLRIFGVLTITVPFVWGLMAL RAWHYDENYSGAIVLWVLLVWGLADSGAYMFGKLFKGLKHLAPKVPSPKQWQFIGG LATAAVISWGYGMWANDVAPVTLTLCISIVAALASVLGDLTSMFKREAGIKDSGLHI

gene
POHGGILDRIDSLTAAPVPVFACLLLLVFRFL"
6745..8097
/gene="yael"
/note="20187"
6745..8097
/gene="yael"
/function="orf; Unknown function"
/note="Residues 1 to 450 of 450 are 100.00 pct identical to residues 1 to 450 of 450 from Escherichia coli K-12 Strain MG1655: B0176"
Strain MG1655: B0176"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG54478.1"
/db_xref="GI:12512901"
/translation="MLSFLWDIASFTVALGVLTIVHEFGHFWARRCGVVERFSTIGF GKALWRRTDKLGTEYVIALPLGGYVKMLDERAEPVVPPELRHFAFNKSVGQRAAIIA AGPVANFIFAI FAYLVLFVIGVPGVPVVEGAANSAIAAQIAPGTGELKAVDGIETP DWDAYRLQLVDKIGDESTITVAPFGSQRRDVKLDLRHWAPEPKDVPVSLGIRPR GPQIEPVLENVQPNASAKAGLQAGDRIVKVDGQPLTQWVTFVMLVRDNPNGKSLALEI

alignment_scores
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block
US-09-701-711-2 x AE005193 ..
Align seg 1/1 to: AE005193 from: 1 to: 10649
648 ProPheTyrGluAsnPhetYrAlaGlyGly 657
|||||
10065 CCGTCTACGAGAACTCTATGCGCGTGT 10094

seq_name: gb_ba:AE004297

seq_documentation_block
LOCUS AE004297 10901 bp DNA linear BCT 31-JUL-2000
DEFINITION Vibrio cholerae chromosome I, section 205 of 251 of the complete chromosome.
ACCESSION AE004297 AE003852
VERSION AE004297.1 GI:9656810
KEYWORDS
SOURCE Vibrio cholerae.
ORGANISM Vibrio cholerae
REFERENCE
1 (bases 1 to 10901)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickley, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C. and White, O.
DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae
JOURNAL Nature 406 (6795), 477-483 (2000)
MEDLINE 20406833
REFERENCE
2 (bases 1 to 10901)
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickley, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragoi, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Nierman, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.
Direct Submission
JOURNAL Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1..10901
/organism="Vibrio cholerae"
/strain="N16961"

```
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/notes="biotype: El Tor"
/complement(69. .1124)
/gene="VC2250"
/complement(69. .1124)
/gene="VC2250"
/notes="similar to GB:X54797 SP:P21645 PID:1208950
PID:41470 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="UDP-3-O-3-hydroxymyristoyl glucosamine
N-acyltransferase"
/protein_id="AAF95394.1"
/db_xref="GI:9656811"
/translation="MTTLTLELATITGGELFGDPALVSAPMPKQALGHVTFILSN
PKYSKHLGDKATVIMKVTSERELCPNSALVADPYFAFKAQAALDSTPSPAHGIAP
SAVIAEDAKLGHVSGANAVIESGVQLGDNVIGAGCFIGKQARLGNITKLANVTI
YHKVEIGSDCLIQSGTVIGADGFGYANERGEWIKIPQLGSVRIGDRVEIGACTTIDRG
ALDDTVIEDNVIIDNLOLQIAHNHVICYSALAGCTVIAGSTRIGKYCIIGASVINGH
IETADGVITIGMGWRSIEEKGMYSGGIPLQTNKEWRKTAARVHRIDDMHRRKLKALE
KLEQSDTVQPDNSQAE"
/complement(1134. .1643)
/gene="VC2251"
/complement(1134. .1643)
/gene="VC2251"
/notes="similar to PID:1935023; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="outer membrane protein OmpH"
/protein_id="AAF95395.1"
/db_xref="GI:9656812"
/translation="MKMKIAASLGILIISSSMANABAAQKIGVINTAQVQALPQ
REVVLQKMEEFKDAELQAIQADAKRIEKLROGLMQDQDEVEKLRIETIGQLDSK
YKIAQALQASARAEAEKQKLFKVIQDAVKVAEKEGYDIVLDTSSMQYGRKEHNL
SEKVIKAIK"
/complement(1658. .4069)
/gene="VC2252"
/complement(1658. .4069)
/gene="VC2252"
/notes="similar to GB:U00096 SP:P39170 PID:1552754
PID:1786374; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="surface antigen"
/protein_id="AAF95396.1"
/db_xref="GI:9656813"
/translation="MAMKOILLATLLATSANGAEKFFVVQDQIDGLORVALGALL
KMPVRVGSVDQVANIILKALYSSGNFEDVKVLRDGNILMVQKERPTIASVSFSGN
KAKEBQLQKNEASIRVEALDRITLSNIEKLEDFFIYSGKYNATKAVVTPLPKR
NRADLKFFTEGSAKIQIINFIGNQVDSDELLSRFNLDNVAMWNLADDKYKQV
LAGDIBALTYLDRLGYKFDVSTQVAISPDKGVYITLNLNEGSPYTVSKVQFGE
LMGKEAFTSLIPFEIGETNGSAVRLSESKVLGSGYAYPOVRTIPEFDDEKQO
VSLVHVHAGKGVVVDIRFVGNNSRDEVLREMPQESWLNKSDIETGKTRNLRL
GFTEFVEQTVRPGSDQVDLYSVKEANSNVNFGVIGESGVSVQVGLQDNNFL
GSGNFGVAMINDYOKNLTLEDRPYNLNDGVSGLGVKVFYQNFSEAGIYDYNES
YDTSLTWGPFPDLNRFEGIGYTHNKIGNLTPYLQVENFLAAQNSIDSGNLLTDD
FDINLSWNTLNNSVFFPAGNHORAFYKMTVPGSDAQVFKLQYDVRQYFPLTKHEF
TLLRLGRLYGGNGYOTDKDNLFPFENFYAGGFTSLRFGSGNSAGKAVRDYSGS
NNGSDPATDSDGGNIALASVELIYPTPFASEEARNQIRTSIFIDMASVWDEFDYR
GRADYGNQYIYDSDPTNRTSSYGVALQWVSPMGPLVFLSLAKPIPKIKKDGDEEFFFTT
IGRTF"
/complement(4118. .5476)
/gene="VC2253"
/complement(4118. .5476)
/gene="VC2253"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
```

```
/protein_id="AAF95397.1"
/db_xref="GI:9656814"
/translation="MDLILWNFIATIIALGILVAVHEFGHFWARRCGVKVEFSIOF
GKSIWKRVGHDGTEYSIMPLGGYVKMLDGRVDDVPAAEQAMAFDQSLWKRSAIYS
AGPIFNFLFAIFAYWLVTGTPAVKPVIGEVTPYSTAAQAGLEPGMEIKAVSGVNTP
DWESVNMGLIGHICDDSMITIVSSAEGVGLNEIKTILRDWNEFDPESEKALGFRP
FTPEISNQLTNVSAQAGERAGLVQGTVLQINGQAVQVYVNAIQSHPNAPIAVM
VERAQOQVELTLIPDSRELQSGKVGIPAGIAPKVAEPONVREPLQGFVPESLGKAVE
KSGQVLDLTVSMKLLKLLYGVGLNNLSGPISIAGAGTTADYGFYPLGLALISINL
GIINLPLPMLDGGHLLFFMIEAVIRRPVEKVOEMGIRIGGAILFISLMAVAIFNDET
RL"
/complement(5473. .6681)
/gene="VC2254"
/complement(5473. .6681)
/gene="VC2254"
/notes="similar to SP:P45568 GB:U00096 PID:1552750
PID:1786369; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="1-deoxy-D-xylulose 5-phosphate reductoisomerase"
/protein_id="AAF95398.1"
/db_xref="GI:9656815"
/translation="MRKLTILGATGSGTSLKVIAQNPOQFSIVALVAGVNVAKMVO
LCOQWRPKYAVMATASAASELQGLLKNQAMATEVLYGEEAMCQVAALDDVDVTMAAIV
GAAGLLPTMAAVKAGKRVLLANKEALVMSQGLFIDAVAQSGAELMPVDSERHNAIFQCL
PTEIOTQLGRCDLSOHCIDHILLTGSGGPFYSDLATLDSVTPEQATAHPNWSMGPKI
SVDSATMMNKGLVYIEAKWLFNTSRLOKVLIIHQSVIHSVMQYQDGSVTAQLGEPDM
ATPISYAMAYPERVTAGVPALDFTRLQQLTFMEVDFAFYPCQLAMDACFLGQHATTS
LNAANEVAVDAFLKRIKIRFTDIALINDQVLSKVCATNTQLHCRDLESLELDTMARHF
AHQVILKERQA"
/complement(6730. .7572)
/gene="VC2255"
/complement(6730. .7572)
/gene="VC2255"
/notes="similar to GB:L42023 SP:P44937 PID:1006039
PID:1221024 PID:1205166; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphatidate cytidylyltransferase"
/protein_id="AAF95399.1"
/db_xref="GI:9656816"
/translation="MKQRIITALILAPVLIGILYLPFAFMFLALAVVTLIGFWEWTQ
FVNQSRMLAMIPALLVGGISVALIDFQFPAISNMNTAHFVILGIGSLWLVSSGLAI
TYDRSPLEHSTVRLHGLFTLLPFNSVLELRADTVLSLDPLYGAKLVFLVCFLVW
AADSGAYFVGKSLGCKHMAPSPFNKTIEGLVGVITAMLVGIWVAECFGIQPSMPV
MLLIILLTVISVLGDLVESMPKRVSGIKDSSNIIIPGHGILDRIDSLELTAAPVPVALL
YFLF"
/complement(7590. .8360)
/gene="VC2256"
/complement(7590. .8360)
/gene="VC2256"
/notes="similar to GB:L42023 SP:P44938 PID:1006041
PID:1221025 PID:1205167; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="undecaprenyl diphosphate synthase"
/protein_id="AAF95400.1"
/db_xref="GI:9656817"
/translation="MFVFLMQLSLSDQVLPKHIAVIMDNGRWAKAOCKPRVFGHK
NGVAAVRKISTISARLGIKAVTLPAFSENWRPREEVGLMELPHTVLTSTELKKLHK
NLNRKRVIGDSKSRFSERLQKTAQASQATNSMTGMVYNIAANTGGWDIIQATQALAE
KVTQGEQLQSDIDEVFKQHLTMADLPDVLDIRITSGEGRISNFMFLWQLAYAEYFTF
VWPFEDENCLIEAVTFWVNRERRRFGCTGQVKALMEN"
/complement(8434. .8991)
/gene="VC2257"
/complement(8434. .8991)
/gene="VC2257"
/notes="similar to GB:L42023 SP:P44307 PID:1005819
PID:1220904 PID:1205056; identified by sequence
similarity; putative"
/codon_start=1
```



```

/gene="XF1039"
/translation="hypothesized protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothesized protein"
/protein_id="AAF83849.1"
/db_xref="GI:9105981"
/translation="MLVNTSVGVNVLFTVQHCLLGLVAEMRCLLAILAWRWILVF
SMGCKYRGIV"
4430..4591
/gene="XF1040"
4430..4591
/gene="XF1040"
/translation="hypothesized protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothesized protein"
/protein_id="AAF83850.1"
/db_xref="GI:9105982"
/translation="MAFTGCAVCGVGTFFNGCVAGVQLGRGLMVCVIEWSVLGRSGFY
HSVSSGWS"
complement(4780..5484)
/gene="XF1041"
complement(4780..5484)
/gene="XF1041"
/translation="similar to SPIP10442 (percent identity: 55 %/query
alignment coverage: 81.6 %/subject alignment coverage:
96.5 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="ribonuclease H11"
/protein_id="AAF83851.1"
/db_xref="GI:9105983"
/translation="MVSQDFTQHLPSNTALVAGVDEAGRPLAGPVVAAVVFDPFSQ
PRINGLDSKQLSPACRERLYAHIVERALAYKVVIDSTQIDTFLNIYQATMLGMRLAV
EGVAHVAKSARIDGNRLPKNLPCPAEALVGGDARDITMAASILAKVTRDRHWLHLE
QYPHYGFDKHKGYGTGAHLAALAEHGFCLERHRSFAPVRKMLTLEAHAKQSTHQHNE
NSTPKAFAFNMLLERDD"
complement(5541..6698)
/gene="XF1042"
complement(5541..6698)
/gene="XF1042"
/translation="similar to SPIP10441 (percent identity: 45 %/query
alignment coverage: 97.4 %/subject alignment coverage:
98.2 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder/Start codon shift: -126"
/codon_start=1
/transl_table=11
/product="lipid A disaccharide synthase"
/protein_id="AAF83852.1"
/db_xref="GI:9105984"
/translation="MIQAPRIALIAEASGDHLAGLIQQLRLRLPTAFVIGIGDDMM
RSARQOTWETSLVMTGLTEVLRHLPLKIRRECKRALAWHPDVLIGDAPDENL
TVRWFQKHRTVHVSVPSINAWREKRAKIGASVDVRLCLFPMPPEPIYARIGIDAR
FVGHPNADPIPTQARTARTALGELLSPVLAVLPGSRHSEISQSGTFLEAAGLS
EHLPGHVVIPANTOCTKPLAEQLSRSLTPVMHSHLDNSARTAMLAADVILVSGT
ATEAMLKLRPMVYAVKAPLTVRYVTKLKLKINRFPALPNTLAGEDLAPETQKDCI
APALCAALDKFKHGKQKVTALQNRYLQLHTQLRRNASTRAAEAITELLPR"
complement(6775..7566)
/gene="XF1043"
complement(6775..7566)
/gene="XF1043"
/translation="similar to SPIP10440 (percent identity: 44 %/query
alignment coverage: 98.5 %/subject alignment coverage:
98.9 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine acyltransferase"
/protein_id="AAF83853.1"

```

```

/db_xref="GI:9105985"
/translation="MNKHASLIHPTAVIAPSATLPDVOIGAFTLIGNDVOIDGTII
GSUCTIHGPRIGRNNRFIGOAAIGGSPQDKKFAGETELLIGNNTIREVTTNRGT
GGGGVTSIGNDNWILAYTHIAHCHVGHHCVFSSNNASLAGHVTVGDMVIFSGSAH
QFCRIGYAFIAGMTLGDVDPFTLIGSTLGRINSGLKRRNFTPERITAIKR
AYRTLYVAGLPALAEAKQVAEQAKDNDIDKELQFIETAQRPLLR"
complement(7588..8070)
/gene="XF1044"
complement(7588..8070)
/gene="XF1044"
/translation="similar to GI11718488 (percent identity: 49 %/query
alignment coverage: 91.2 %/subject alignment coverage:
98.0 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="(3r)-hydroxymyristoyl ACP dehydrase"
/protein_id="AAF83854.1"

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AE003941/rev ..
Align seg 1/1 to reverse of: AE003941 from: 1 to: 11711
368 AspGluValLeuArgArgGluMetArgGln 377
|||||
10531 GATGAAGTACTACGTCGTAATGCGTCAG 10502

seq_name: gb_ba:AE004784
seq_documentation_block:
LOCUS AE004784
DEFINITION Pseudomonas aeruginosa PA01, section 345 of 529 of the complete
genome.
ACCESSION AE004784 AE004091
VERSION AE004784.1 GI:9949799
KEYWORDS
SOURCE
ORGANISM
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 12469)
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,
Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,
Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y.,
Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
Smith, K., Spencer, D., Wong, G.K., Wu, Z. and Paulsen, I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
TITLE
Nature 406 (6799), 959-964 (2000)
JOURNAL
2043737
MEDLINE
2 (bases 1 to 12469)
Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P.,
Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,
Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saler, M.H.,
Hancock, R.E.W., Lory, S. and Olson, M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University Of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
Location/Qualifiers
1..12469
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"

```



```
/db_xref="GI:9949808"
/translation="MKRFLLPALLSALMIAEVAHSAFTVSIDIRVNGLQVRSAGSFAEA
LPLVGETTDDALVQATSLFKTFQFQDIOIGRDNGLVTVTVVERPSISSIEIEGNK
ATSKEDLLKGLQSGLAEGEIQFQATLEGRVRELQRYVQAQRYSAEINAEVPOPRN
RVALKINNEGIVAIASHINVGNTVFSSEDDLLFELATITNWSFFKNDKDKAREKL
KDELRSLRYLLDRGYINNDIAQVTSQVSDTKKHVYITVINEGEKTYITDKVLKGLD
KYPVEEVKRLLVQKQGVFSRKVMTTSDLTIRRLGNEGTYTANVNGVPEAHDDDKTV
SVTFVDPGKRAYVNRINFRNGTKTEDEVLRRMROMEGGWASTYLDGSKARLERLG
TKYKVNVEPAVGTDDQVDVNVSVVEQPSGTSVSGPAQAGLILGGSISONNFLG
TGNKYSIGLITREXOTRYNFGVDPVYVTDVGYSLGYNAYRKYTDDELDVYASVYN
SLGAGMSIGPISESRNLGYLSVQRDQIDTGYTVDEIYDFDKEGDNFTNFKASIG
WSESTLNGKVALTRHQSQTLETTLPLGSDLSFYKIDYRGQVFAPLTDNTYTRHFEEL
GCGDGYTERLPFYENYVAGFNSVRGKSTLGPSTSVARNPDGPMKNGQGPS
KGRYTDPODPAFGDNILITGGAELLPFPVKORQRTLVFWDVGSTFTDTCPTK
TTTNCDIGTGNILASSGVGLTWITALGPLSFLATPIKKPDNAETQVQFSLGQTF"
BASE COUNT      2144 a 3895 c 4089 g 2341 t
ORIGIN

alignment_scores:
  Quality:   10.00      Length:      10
  Ratio:     1.000      Gaps:        0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AE004784/rev ..
Align seg 1/1 to reverse of: AE004784 from: 1 to: 12469

368 AspGluValLeuArgArgGluMetArgGln 377
|||||
11324 GACGAAGTGTGCGCGGAAATGCGCAG 11295

seq_name: gb_ba:AE000127

seq_documentation_block:
LOCUS      AE000127      14877 bp      DNA      linear      BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 17 of 400 of the complete
            genome.
ACCESSION  AE000127 U00096
VERSION    AE000127.1 GI:1786370
KEYWORDS
SOURCE     Escherichia coli K12.
            Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE  1 (bases 1 to 14877)
            Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5331), 1453-1474 (1997)
TITLE      JOURNAL
MEDLINE   97426617
PUBMED    9278503
REFERENCE  2 (bases 1 to 14877)
            Blattner,F.R.
            Direct Submission
            Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            3 (bases 1 to 14877)
            Blattner,F.R.
            Direct Submission
            Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            4 (bases 1 to 14877)
            Plunkett,G. III.
            Direct Submission
            Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
```

```
COMMENT
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, Director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES             Location/Qualifiers
     source            1..14877
                        /organism="Escherichia coli K12"
                        /strain="K12"
                        /sub_strain="MG1655"
                        /db_xref="taxon:83333"
     gene              123..884
                        /gene="yaeS"
                        /note="b0174"
                        123..884
                        /gene="yaeS"
                        /function="orf; Unknown"
                        /note="0253; 61 pct identical to 229 residues of an
                        approx.240 aa hypothetical protein Y920_HAEIN SW: P44938"
                        /codon_start=1
                        /transl_table=11
                        /product="orf, hypothetical protein"
                        /protein_id="AAC73285.1"
                        /db_xref="GI:1786371"
                        /translation="MMLSATQPLSEKLPAGRCRHVAIIMDNGRWAKKQKIRAFGKH
                        AGAKSVRAVSAFNAANGIEALTLYAFSSNNRPAQEVSAIMELFWALDSEVSKLHR
                        HNVRLRIIGDTSRNSRLQERIRKSEALTAGNTGLTNTAANYGGRWIDVQGVQLAE
                        KVOQGNLPQDIDDEMLNQHVCMLAPDLVIRGTGHEIRISNFWLLQIAYAEIYFTD
                        VLWPDFDEQDFEGALNAFANRRRRFGTPEGDETA"
     promoter          870..900
                        /note="factor Sigma70; predicted +1 start at 195687"
     promoter          899..927
                        /note="factor Sigma70; predicted +1 start at 195714"
     promoter          956..987
                        /note="factor Sigma70; predicted +1 start at 195774"
     gene              1005..1754
                        /gene="cdsA"
                        /note="b0175"
                        1005..1754
                        /gene="cdsA"
                        /EC_number="2.7.7.41"
                        /function="enzyme; Fatty acid and phosphatidic acid
                        biosynthesis"
     CDS               /note="0249; 100 pct identical to CDSA_ECOLI SW: P06466"
                        /codon_start=1
                        /transl_table=11
                        /product="CDP-diglyceride synthetase"
                        /protein_id="AAC73286.1"
                        /db_xref="GI:1786372"
                        /translation="MLAAEWGQLSGFTTSQRVWLAVLGLLLALMLFLPEYHRNI
                        HOPLVEISLWASLGWIVALLLVFPGSAAIWRNSKTLRILFGLVLTVPFFWGMAL
```

```
protein_bind
RAWHDYDNHSGAIWLIVYMLVWAGSGAYMFGKLFKHKHKLAPKVPSPGKTWQGFIGG
LATAAVISWGYGMANLDVAPVTLLIGSIVAALASVLGDLTFESMFKEAGIKDGSGLHI
PGHGILDRIDSLTAAPVPFACLLLVFRTL"
1421..1432
/ gene="cdsA"
/ note="No predicted promoter"
/ bound_motety="PutA predicted site"
1766..3118
/ gene="yael"
/ note="b0176"
1766..3118
/ gene="yael"
/ function="orf; Unknown"
/ note="o450; 100 pct identical to fragment YAEI_ECOLI
SW:P37764 but contains 302 additional C-terminal residues"
/ codon_start=1
/ transl_table=11
/ product="orf, hypothetical protein"
/ protein_id="AAC73287.1"
/ db_xref="GI:1786373"
/ translation="MLGFLDLASFIVALGVLIIVVHEFGHFWARRCGVRVERFSIGF
GKALRRDKLGTGVYVIALPLGCVKMLDERAEPVPELRHHAFNNKSVGORAAIIA
AGPVANFIAFALWLVFIIGVGVPRVPGVEIAANSIAAEQIAPGIELKAVDGIETP
DMDAVRQLQVDKIDESTTITVAPFGSDQRDKLDLRHAFEPDKEDPVSLSGIRP
GPOTEPLENVQPNASAKAGLQADRIKRVQDQPLTQWVTFVMLVDNPGKSLALEI
ERQSPSLSTLIPKSGKNGKAIGFVIEPKVIPLPEYKVVQYGVGFNPATVEATDKT
WQMKLVSMGLKLTGDKVGLNNLSGPISTAKGAGTAELGVVYLPFLALISVNLGI
INLPLPLVDGHLLEAIEKIKGPPVSERVQDFCYRIGSGILLVLLMGLALFNDFSRL"
3148..5580
/ gene="yael"
/ note="b0177"
3148..5580
/ gene="yael"
/ function="orf; Unknown"
/ note="o810; 45 pct identical (29 gaps) to 808 residues of
an approx. 800 aa protein D151_HAEIN SW: P46024"
/ codon_start=1
/ transl_table=11
/ product="orf, hypothetical protein"
/ protein_id="AAC73288.1"
/ db_xref="GI:1786374"
/ translation="MAMKLLIASLFSSTVYGAEGFVVKDIHFEGLRVAVGAALL
SMVPTGDTWNDEISITRALFATGFNEDVRVLRDGDITLVQVKERTIASIFSGN
KSVKDDMLKNLEASGVVGVESLDRITADIEKLEDFYYSVGKYSASVAVVPLPR
NRVLDKLVQFQGVSAEQIQINIVGNHAFITDELISHFQLRDEVPWNVGDRKYQOK
LAGLETLRSYVLDRGVARENIDSTQVSLPDKGIYVYNTIPEGDQYKLSGVESGN
LAGHSAIEQLTKPELGELNGTKVIMEDDIIKLLGRYGTAYPRVSPMEINDADKT
VKLRVNDAGNRFTYVRKIRFEGNDTSKDALRREMOMEGAWGLSDLDQGERLNL
GFFETVDTQRPVQSDVDVYKVKERNITGSFNGIGYTESGVSFQAGVQKQNL
GTGVAVNGTKNDYQYAEISVNTPYFTVDGSLGRLEFVNDQADADLSQDYNKS
YGTDTVLGFPINEYNSLRAGLVYVHNSLNMPQVAMRVLYLSMGEHPSTSDQNSFK
TDDTFNYGWTYKLDKRGYPTDGSRYNLGKVTIPGSDNEYKYKTLDTATVPIDDD
HKWVYLRTRWGYDGLGKEMPEYFNFYAGGSTVVRGFSNTIGKRAVTFPHQASNY
DPDYDECATQDGAIDCKSDDAVGNAMAVASLEFITPTFFISDKYANSVRSFFWD
MGTVNDWDSQYSGYPYSDPSNIRMSAGIALQWMSPLGLVFSYQAQPFKKYDGGK
AEQFQFNIGKTW"
5583..5611
/ note="factor Sigma70; predicted +1 start at 200398"
5702..6187
/ gene="hlpA"
/ note="b0178"
5702..6187
/ gene="hlpA"
/ function="factor; Basic proteins - synthesis,
modification"
/ note="o161; 100 pct identical to HLP_A_ECOLI SW:
P11457; similar to ompH of other bacterial species"
/ codon_start=1
/ transl_table=11
/ product="histone-like protein, located in outer membrane
or nucleoid"
/ protein_id="AAC73289.1"
```

```
/db_xref="GI:1786375"
/ translation="MKKWLALQAGLGLATSAQADKIAIVNMGSLFQQVAGKTGVSN
TLENEFGRASELQRMETDLOAKMKKQSMRAGSDRTKLEKDVMAOROTFAQKAQAFE
QDRARRSNEERGKLVTRIQATVAKSVANSQDIDLVDANAVAYNSSDVKDITADVLKQV
K"
6191..7216
/ gene="lpxD"
/ note="b0179"
6191..7216
/ gene="lpxD"
/ EC_number="2.3.1.-"
/ function="enzyme; Surface polysaccharides and antigens"
/ note="o341; 100 pct identical to LPXD_ECOLI SW: P21645
but includes initiator met"
/ codon_start=1
/ transl_table=11

alignment_scores
  Quality: 10.00 Length: 10
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block
US-09-701-711-2 x AE000127 ..
Align seg 1/1 to: AE000127 from: 1 to: 14877
648 PropHetYrGluAsnPhetYrAlaGlyGly 657
|||||
5086 CCGTCTACGAGAACTCTATGCCGTGGT 5115

seq_name: gb_ba:AE008705

seq_documentation_block
LOCUS AE008705 22997 bp DNA linear BCT 25-OCT-2001
DEFINITION Salmonella typhimurium LT2, section 13 of 224 of the complete
genome.
ACCESSION AE008705 AE006468
VERSION AE008705.1 GI:16418723
KEYWORDS
SOURCE
  Salmonella typhimurium LT2.
  Salmonella typhimurium LT2
  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Salmonella.
REFERENCE
  1 (bases 1 to 22997)
  McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
  Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
  Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
  Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Flores,L., Miller,W.,
  Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  Complete genome sequence of Salmonella enterica serovar Typhimurium
  LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
PUBMED 11677609
REFERENCE 2 (bases 1 to 22997)
  The Salmonella typhimurium Genome Sequencing Project.
AUTHORS Direct Submission
TITLE Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/geneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
and Pedro Romero and Peter Karp at EcoCyc;
http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites
```

were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

FEATURES	source
	location/Qualifiers
	1. .22997
	/organism="Salmonella typhimurium LT2"
	/strain="LT2; SGSC 1412; ATCC 700720"
	/db_xref="ATCC:700720"
	/db_xref="taxon:99287"
	98. .103
	/gene="frr"
	/note="Putative RBS for frr; RegulonDB:STMS1H000586"
	98. .668
	/gene="frr"
	/note="STM0219"
	111. .668
	/gene="frr"
	/note="Ortholog of E. coli ribosome releasing factor (AAC73283.1); Blastp hit to AAC73283.1 (185 aa), 93% identity in aa 1 - 185"
	/codon_start=1
	/transl_table=11
	/product="ribosome releasing factor"
	/protein_id="AAL19183.1"
	/db_xref="GI:16418724"
	/translation="MISDIRDAEVRMEKVEAFKTIQISKVRTGRASPSLDGIWVEY YGPTPLQRLASVTVEDSRTLKINVEDRSMGPAVEKAIMASDLGNPSSAGTDIRVPL PLTEERRKDLTKIVRGEAQARVAVNRVRDANKVKALLKKAISEDDDRQSEEV QKMTDAIKKVAALADKAEALMQF"
	798. .2005
	/gene="dxr"
	/note="STM0220"
	798. .803
	/gene="dxr"
	/note="Putative RBS for dxr; RegulonDB:STMS1H000587"
	809. .2005
	/gene="dxr"
	/EC_number="1.1.1.-"
	/note="Ortholog of E. coli putative ATP-binding component of a transport system (AAC73284.1); Blastp hit to AAC73284.1 (398 aa), 88% identity in aa 1 - 397"
	/codon_start=1
	/transl_table=11
	/product="1-deoxy-D-xylulose 5-phosphate reductoisomerase"
	/protein_id="AAL19184.1"
	/db_xref="GI:16418725"
	/translation="MKQTLILGSTIGCTSLDVRHNPDSPFVIALVACKNVMAR EOCLESPRYAVMDTSSAEQKIMLQHQSGRTVLGQQAAACEMALDEYGHYMAIV GAAGLPLTAAIRAGRTILIANKESLVTGRLFMDEVKRSNARLLPVDSHNNAIFQSL PSQIOHNLVADLEQNGVTIILTGSGGPPRETQPCDILAAMTPDOACRHPNWSMGRI KIVSATMKNGLVEYIARWLFNASARQMEVLIHQSVIHSWVYQGSVLQALGEPDM RTPIATMAMPNVRTSGAQLDFCKLSALTFSAPDYQRYPCLKLANEAFQGGQAATA LNAANEITVAALQAQQIRFTDIAGLNLAVLERMDLQEPASVEDVLQVDATAREVARRQ VIRLSR"
	2299. .2304
	/gene="upps"
	/note="Putative RBS for upps; RegulonDB:STMS1H000588"
	2299. .3076
	/gene="upps"
	/note="STM0221"
	2318. .3076
	/gene="upps"
	/EC_number="2.5.1.31"
	/note="Ortholog of E. coli orf, hypothetical protein (AAC73285.1); Blastp hit to AAC73285.1 (253 aa), 93% identity in aa 2 - 253"

	/codon_start=1
	/transl_table=11
	/product="undecaprenyl pyrophosphate synthetase (dl-trans,poly-cis-decaprenylcistransferase)"
	/protein_id="AAL19185.1"
	/db_xref="GI:16418726"
	/translation="MLSATOPVSENLPAGHCRHVAIIMDGNRWAKKQGIKRAFQHKH GAKSVRRVAFSAANGIDALTLYAFSSNNRNPQAEYSALMELFWALDSVKSILHRH NVRLRIIGDISRFNSRLQERIKSEALTHTNTGLTATANYGGRWDIVQVRLAEQ VQGLVLRPDQIDEERLGGQICMHELAPVDLVIRTGGEHRISNLLWQIAYAEILFTDV LMPDFDEQFEGLAHAFANKRERFEGGTEPGDDKA"
	3076. .3081
	/gene="cdsA"
	/note="Putative RBS for cdsA; RegulonDB:STMS1H000589"
	3076. .3946
	/gene="cdsA"
	/note="STM0222"
	3089. .3946
	/gene="cdsA"
	/EC_number="2.7.7.41"
	/note="Ortholog of E. coli CDP-diglyceride synthetase (AAC73286.1); Blastp hit to AAC73286.1 (249 aa), 95% identity in aa 1 - 249"
	/codon_start=1
	/transl_table=11
	/product="CDP-diglyceride synthase"
	/protein_id="AAL19186.1"
	/db_xref="GI:16418727"
	/translation="MLKYLGISAFVLIPAVTAAFLPLPPVGFALTITLVCMALAWENG QLSGFAARSQRVWLAVLCGLLALLFLPEYHHNIROPVLEMSLWASLGWVVALLL LVFYPGSAIWRNSKTLRILFGLTIVPFFWGLALRAWHYDHENYSIAIWLIVMIL VWGADGAYMFGKLFKHKLAPSPGKTQOFGIGGLATAAIVISWGYGMANLVAPV ILLICSVVAALASVLGDLTSMFPKREAGINDSGHLIPGHGGLIDRIDSITAAVPVFC LILLVFPRL"
	3950. .5310
	/gene="yael"
	/note="STM0223"
	3950. .3955
	/gene="yael"
	/note="Putative RBS for yael; RegulonDB:STMS1H000590"
	3958. .5310
	/gene="yael"
	/note="Ortholog of E. coli orf, hypothetical protein (AAC73287.1); Blastp hit to AAC73287.1 (450 aa), 88% identity in aa 1 - 450"
	/codon_start=1
	/transl_table=11
	/product="putative membrane-associated Zn-dependent protease"
	/protein_id="AAL19187.1"
	/db_xref="GI:16418728"
	/translation="MLSILMNLAFFIITGVLITVHEFGHEFWARRCGVRVERSIGF GKALWRTDRYGTVEYVIALIPGGVVKMLDERAEPVAPELRRHAFNKKTVQORAAIIA AGVPANIFAIFAYVWVFIIGVGVPRVIGEITPNSIAAQAIAPGTELKAVDGIEPT DPAIVRLQVSKSGDQQTTSVAPFGSDQKQDRLDLRHNWAFEPKDPVSSGLIRPR GWQVRLVSEVQANSAASKAGLQAGDRIVKVDGQPLTKQMMKFVTRNDPNKGLALEI EROGSALSLTLPTDKTKNGKAGFAGVVPKIIPLEPYKTIROYGPSFSAILEATDKT WQMLKVSMLGKITGDVKNLNSGPISTAOAGMSEFGVIVYLMFLALISVNLGI INLPLPVLGGHLLFLAIEKLGKGPVSERVQDFSYRIGSILLVLLMGLALFNDFSRL "
	5342. .7756
	/gene="yael"
	/note="STM0224"
	5342. .7756
	/gene="yael"
	/note="Ortholog of E. coli orf, hypothetical protein (AAC73288.1); Blastp hit to AAC73288.1 (810 aa), 92% identity in aa 1 - 810"
	/codon_start=1
	/transl_table=11
	/product="putative outer membrane antigen"
	/protein_id="AAL19188.1"
	/db_xref="GI:16418729"

/translation="MAMKLLIASLLFSSATVYGAEGFVVVKDIHFEGLRQVAVGAALL
SPVYGDVTDNEDISITRALFATGNEFDVRLDGNLTLLQVQKRPITIASITFSGN
KSVYKQDLQNLAEASGVYGESLDRITLSDIEKEDFYISVGKYSASVAVVTPLR
NRVDLKLVEQGVSAKIQIINIVGNHAFSTEELISHFOLRDEVPWNVVGVDRKYOKQ
LAGDLTLRSYLLDRGYARFNIDSTQVSLTPDKKGIYIVNITEGDOYKLSGVQVSGN
LAGHSAEINLAKIEPGELYNGKVTKMEDDIKKLLRGYIAYPRVQSQPEINDADKT
VLRNVNAGNFYVRKIRFENDTSKSVLRREMGAWGLSDLVQDGRKLNRL
GFFETVDTDPGSPQDVVYKVKERTSGNFNGIGYTESGVFOGVQDQNL
GTGYSVINGTKNDYQYSELSVTNPFYVDVGLGRIFYNDFOADADLSDYTNKS
YGDVTLGFPINEYTLNLAGLVVHKNLSMOPQLAMDRYLESMQSDATSSFAADDF
TENYCWYKLDRGVEPTDGSRVNLTKGVTIFGSDNEYKYSILDTATYVPIDNHKKW
VLGRHWGSDGLGKEMPEFYENFAGSSVTRGFSQNTIGPKAVYKNGAHTSWDDND
DYEDCTQESGCKSDDAVGNAMAVASLEFITTPTISEKYANSVRTSFFWMDGTWMDT
NNDPSAPSDVPDYSDPGNIRMSAGIALQWMSPLGPFVSYAQPFKKYDCGKAEQFOF
NICKTW"

alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AE008705

Align seg 1/1 to: AE008705 from: 1 to: 22997

648 ProPheTyrGluAsnPheTyrAlaGlyGly 657

|||||
7268 CCGTTCATGAGAACTCTATGCCGGTGT 7297

seq_name: gb_htg:AC104418

seq_documentation_block:

LOCUS AC104418 62088 bp DNA linear HTG 10-DEC-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-1113M11 map 15, LOW-PASS
SEQUENCE SAMPLING.

ACCESSION AC104418

VERSION AC104418.1 GI:17432864

KEYWORDS HTG; HTGS_PHASE0.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 62088)

Birren.B., Linton.L., Nusbaum.C. and Lander.E.

Unpublished

REFERENCE 2 (bases 1 to 62088)

Birren.B., Linton.L., Nusbaum.C., Lander.E., Ali.A., Allen.N.,
Anderson.S., Barina.N., Bastien.V., Boguslavsky.L., Boukhgalter.B.,
Brown.A., Camarata.J., Campopiano.A., Chang.J., Chazaro.B.,
Choepe.I., Colangelo.M., Collins.S., Collins.S., Cook.A.,
Cooke.P., Dearellano.K., Dewar.K., Diaz.J.S., Dodge.S., Faro.S.,
Ferreira.P., FitzHugh.W., Gage.D., Galagan.J., Gardyna.S.,
Ginde.S., Gord.S., Goyette.M., Graham.L., Grand-Pierre.N.,
Hagos.B., Hearford.A., Horton.L., Hulme.W., Illiev.I., Johnson.R.,
Jones.C., Kamat.A., Karatas.A., Kells.C., LaRoque.K.,
Lamarez.R., Landers.T., Lehoczy.J., Levine.R., Liu.G.,
MacLean.C., Macdonald.P., Major.J., Marquis.N., Matthews.C.,
McCarthy.M., McEwan.P., McKernan.K., McPheeters.R., Meldrim.J.,
Meneus.L., Mihova.T., Mlenga.V., Murphy.T., Naylor.J., Nguyen.C.,
Norbu.C., Norman.C.H., O'Connor.T., O'Donnell.P., O'Neil.D.,
Olliver.J., Peterson.K., Phunkhang.P., Pierre.N., Pollara.V.,
Raymond.C., Retta.R., Rieback.M., Riley.R., Rise.C., Rogov.P.,
Roman.J., Rosetti.M., Roy.A., Santos.R., Schauer.S., Schupback.R.,
Seaman.S., Severy.P., Spencer.B., Stange-Thomann.N., Stojanovic.N.,
Strauss.N., Subramanian.A., Talamas.J., Tesfaye.S., Theodore.J.,
Topham.K., Travers.M., Travis.N., Trigilio.J., Vassiliev.H.,
Viel.R., Vo.A., Wilson.B., Wu.X., Wyman.D., Ye.W.J., Young.G.,
Zainoun.J., Zembek.L., Zimmer.A. and Zody.M.

Direct Submission

Submitted (10-DEC-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21692
Center clone name: 1113_M_11

* NOTE: This record contains 79 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 677: contig of 677 bp in length
678 777: gap of 100 bp
778 1456: contig of 679 bp in length
1457 1556: gap of 100 bp
1557 2244: contig of 688 bp in length
2245 2344: gap of 100 bp
2345 3033: contig of 689 bp in length
3034 3133: gap of 100 bp
3134 3843: contig of 710 bp in length
3844 3943: gap of 100 bp
3944 4634: contig of 691 bp in length
4635 4734: gap of 100 bp
4735 5383: contig of 649 bp in length
5384 5483: gap of 100 bp
5484 6148: contig of 665 bp in length
6149 6248: gap of 100 bp
6249 6928: contig of 680 bp in length
6929 7028: gap of 100 bp
7029 7734: contig of 706 bp in length
7735 7834: gap of 100 bp
7835 8519: contig of 685 bp in length
8520 8619: gap of 100 bp
8620 9294: contig of 675 bp in length
9295 9394: gap of 100 bp
9395 10093: contig of 699 bp in length
10094 10193: gap of 100 bp
10194 10875: contig of 682 bp in length
10876 10975: gap of 100 bp
10976 11675: contig of 700 bp in length
11676 11775: gap of 100 bp
11776 12448: contig of 673 bp in length
12449 12548: gap of 100 bp
12549 13247: contig of 699 bp in length
13248 13347: gap of 100 bp
13348 14027: contig of 680 bp in length
14028 14127: gap of 100 bp
14128 14828: contig of 701 bp in length
14829 14928: gap of 100 bp
14929 15623: contig of 695 bp in length
15624 15723: gap of 100 bp
15724 16413: contig of 690 bp in length
16414 16513: gap of 100 bp
16514 17205: contig of 692 bp in length
17206 17305: gap of 100 bp
17306 17994: contig of 689 bp in length
17995 18094: gap of 100 bp
18095 18766: contig of 672 bp in length
18767 18866: gap of 100 bp
18867 19576: contig of 710 bp in length
19577 19676: gap of 100 bp

```
* 19677 20364: contig of 688 bp in length
* 20365 20464: gap of 100 bp
* 20465 21162: contig of 698 bp in length
* 21163 21262: gap of 100 bp
* 21263 21953: contig of 690 bp in length
* 21953 22052: gap of 100 bp
* 22053 22740: contig of 688 bp in length
* 22741 22840: gap of 100 bp
* 22841 23539: contig of 699 bp in length
* 23540 23639: gap of 100 bp
* 23640 24346: contig of 707 bp in length
* 24347 24446: gap of 100 bp
* 24447 25138: contig of 692 bp in length
* 25139 25238: gap of 100 bp
* 25239 25909: contig of 671 bp in length
* 25910 26009: gap of 100 bp
* 26010 26694: contig of 685 bp in length
* 26695 26794: gap of 100 bp
* 26795 27482: contig of 688 bp in length
* 27483 27582: gap of 100 bp
* 27583 28279: contig of 697 bp in length
* 28280 28379: gap of 100 bp
* 28380 29076: contig of 697 bp in length
* 29077 29176: gap of 100 bp
* 29177 29846: contig of 670 bp in length
* 29847 29946: gap of 100 bp
* 29947 30618: contig of 672 bp in length
* 30619 30718: gap of 100 bp
* 30719 31410: contig of 692 bp in length
* 31411 31510: gap of 100 bp
* 31511 32184: contig of 674 bp in length
* 32185 32284: gap of 100 bp
* 32285 32961: contig of 677 bp in length
* 32962 33061: gap of 100 bp
* 33062 33738: contig of 677 bp in length
* 33739 33838: gap of 100 bp
* 33839 34545: contig of 707 bp in length
* 34546 34645: gap of 100 bp
* 34646 35343: contig of 698 bp in length
* 35344 35443: gap of 100 bp
* 35444 36096: contig of 653 bp in length
* 36097 36196: gap of 100 bp
* 36197 36892: contig of 696 bp in length
* 36893 36992: gap of 100 bp
* 36993 37678: contig of 686 bp in length
* 37679 37778: gap of 100 bp
* 37779 38459: contig of 681 bp in length
* 38460 38559: gap of 100 bp
* 38560 39264: contig of 705 bp in length
* 39265 39364: gap of 100 bp
* 39365 40038: contig of 674 bp in length
* 40039 40138: gap of 100 bp
* 40139 40809: contig of 671 bp in length
* 40810 40909: gap of 100 bp
* 40910 41593: contig of 684 bp in length
* 41594 41693: gap of 100 bp
* 41694 42373: contig of 680 bp in length
* 42374 42473: gap of 100 bp
* 42474 43168: contig of 695 bp in length
* 43169 43268: gap of 100 bp
* 43269 43966: contig of 698 bp in length
* 43967 44066: gap of 100 bp
* 44067 44761: contig of 695 bp in length
* 44762 44861: gap of 100 bp
* 44862 45541: contig of 680 bp in length
* 45542 45641: gap of 100 bp
* 45642 46349: contig of 708 bp in length
* 46350 46449: gap of 100 bp
* 46450 47138: contig of 689 bp in length
* 47139 47238: gap of 100 bp
* 47239 47939: contig of 701 bp in length
* 47940 48039: gap of 100 bp
* 48040 48697: contig of 658 bp in length
```

```
* 48698 48797: gap of 100 bp
* 48798 49471: contig of 674 bp in length
* 49472 49571: gap of 100 bp
* 49572 50253: contig of 684 bp in length
* 50256 50355: gap of 100 bp
* 50356 51055: contig of 700 bp in length
* 51056 51155: gap of 100 bp
* 51156 51844: contig of 689 bp in length
* 51845 51944: gap of 100 bp
* 51945 52626: contig of 682 bp in length
* 52627 52726: gap of 100 bp
* 52727 53401: contig of 675 bp in length
* 53402 53501: gap of 100 bp
* 53502 54198: contig of 697 bp in length

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AC104418 ..
Align seg 1/1 to: AC104418 from: 1 to: 62088
430 GlySerSerThrIleAlaAlaGlyTyrSer 439
|||||
21572 GGCAGCTCCACTATAGCAGCGGGTACAGC 21601

seq_name: gb_htg:AC017607

seq_documentation_block:
LOCUS AC017607 81255 bp DNA linear HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017607
VERSION AC017607.1 GI:6554390
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 81255)
AUTHORS Adams, M. and Venter, J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211447 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
  source
    1..81255
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
BASE COUNT 22419 a 18447 c 18706 g 21683 t
ORIGIN

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AC017607//rev ..
Align seg 1/1 to reverse of: AC017607 from: 1 to: 81255
789 SerTyrAlaLysProLeuAsnLysGln 798
```



```

/feature="fhuD"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ferrichrome-iron transport protein FhuD"
/protein_id="AAB08582.1"
/db_xref="GI:1552730"
/translation="MSGPLISRRLLTAMALSPLLMQMNTAHAAIDPNRIVALEWL
PYELLALGIVGYVADTINYRLWSEPPDSDVIDGLRTEPNLELLETKPSPFVW
SAGYGPSPMLARIAPGRGNFSDGKQPLAMARKSLTEADLLNLQSAATHLAQYED
FTRMKPRVKKRGAPLLTLLIDPRHMLVFGPNSLFOELDEYGIIPNAWQGETNFWG
STAVSDRLAAVKDQVLCFCFHDNSKMDALMATPLWQAMPFVRAGRFRVPAVFEY
ATLSAMFVRVLDNALGKA"
2537..4519
/feature="fhuB"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="ferrichrome-iron transport protein FhuB"
/protein_id="AAB08583.1"
/db_xref="GI:1552731"
/translation="MSKRIALPALLLIVIVATLTMNESQALPRSQWAAQANSP
DIDVLEQMFHYSILPRALISLLYGAGLGVLFQOVLRLNPLAEPTTLGVATGAQUG
IYVITWALPAGWASQFAQAGACVVLGVFGVANGKRLSPVTLILAGLVSLYCGAI
NOLLVIFHDDQOSMFLSTGTQTQDGGVERLPQLLGVMLTLLRLPTLMGLD
DGVANGLASLARLAALSLAIVISALLVNAVGITGIFAPLAKMLGARLLPR
LMASLICALIWLSDQITLWLRVMEVSTGVTALIGAPLLMLPLRLRISAPDM
KYNRAAERQHVAFALAGVLLMAMVVALSFGDAGHGTWASGALLEDLMPWRP
RIMAAFAQVHLAVAGCIQRLTGMPMASPEVLGISSAGFAQVGLMLFLVFNAGFWL
LPAGSLGAATVLLIIMIAAGRGFSFHRMLLAGMALSTLMLLMQLQASGDPRAQWL
LFWISGNTNADQVWRTGIVMVLITLPLCRWLTILPLGGDTARAVGMLTPTR
IALLLAACLATATMTIGPLSFVGLMAPHIAARMGFRRTMPEHIVISALVGLLILVFA
DWCGRMVFPPQIPAGLLSTFGAPYFIYLLKQSR"
complement(4677..5957)
/feature="gsa"
/complement(4677..5957)
/feature="gsa"
/codon_start=1
/transl_table=11
/product="glutamine-1-semialdehyde aminotransferase"
/protein_id="AAB08584.1"
/db_xref="GI:1552732"
/translation="MSKENLYSARELIPGVNSPVRFTGVTGPTLFIKADGAYL
YDVGKATIDYVSGMPVGLHNPAPIRNAVIEAAREGLSFGAPTEMYKNAQLYTEL
VPTDMVRVNSGTATSAIRLARGFTGRDKIIFEGCYHGHADCLLVKAGSGALT
GQPNSPGPVADPAKYTLACTVTNDLASVRAAPQYPOETACIIVEPVAGNMNCVPLPE
FLPGLRLCDEFGALLIIDEVMTGPRVALAGADYGVVVDLTCLGKIIGGMPVCAF
GGRDVMALAPGVPYQAGTSLGNPIAMAGFACLNEVAQPGVHETIDELTRLAEG
LLEAAERAGIPLVNVHVGGMFGIFFTDAESVTCYQDVMACDVERKRFHMMLDGCV
LAPSAFEAGFMSVAHSMEDINNTIDAARRVFAKL"
6182..7603
/feature="similar to voltage-gated chloride channel protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08585.1"
/db_xref="GI:1552733"
/translation="MKTDTPSLETPOAARLRRLIROLLRDKTPTLALFMAAVVGT
LVGLAAVDFGVAWLQNRMCALVHTADNYPDLLTVAFLCSAVLAMEGYFLVRYP
EAGSGSIEGALDQRPVRWRVLPVKFFGGTLGGWVGLRGEPVQVIGNGR
MYLDFRLKDEARHTLTAAGAAALFAPNAPLAGILFIEEMSPRYTLISIKAV
FIGVMTIMYRINHEVALIDVGLSDAPLNTLWLYLILGIIFIGIPFNKWLGM
QDLHRVHGNTITKWLMMGAGTGLLGVFAPATSGGFNLIPATVAGFNKMLV
FIFVARVITLILCFSSGAPGGIFAPMLALGTVLGTFAGWAVELFPQVHLEAGTFAIA
GMGALLAASIRAPLTGTIILVLEMTDNYQLILPMIITGLGATLLAQFTGGKPLYSAILA
RTLAQAEQRLARSKAASENT"
7685..8029
/feature="hypothetical protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental

```

```

/protein_id="AAB08586.1"
/db_xref="GI:1552734"
/translation="MSDDVALPLEFTDAANKVKSIADEDNENLKLRYVITGGCGSG
FOYGTTFDDQVNEGDMTIEKQGVGLVDPMSLQYLVGGSDYDTEGLEGSRFIYNPNA
KSTCGGSSFS"
complement(8076..8699)
/feature="similar to Vibrio cholerae RecA protein"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08587.1"
/db_xref="GI:1552735"
/translation="MLVYWLIDIVGTAVFAISGVLLAGKLRMDPFGVLVGLWVAVGGG
TIRDMALDHGPVFWKDPDLDLVVAMVTSMLTIVLVRPRLPKWMLPDLVAGLAVFV
GIGVNAKAFNAEAGPLIACMGVITGVGGIIRDLAREIPMLIRTIYATACIIGIV
HATAYTFSVPLETASMMGMVVTLLIRLAIRAHWLKLPFDALDENG"
complement(8737..9537)
/feature="yadT"
/complement(8737..9537)
/feature="yadT"
/feature="hypothetical"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08588.1"
/db_xref="GI:1552736"
/translation="MAKSLFRALVALSFLAPLWLNAAPRVITLSPANTELAFAAGITP
VGVSSYDPPQAKIEQVSTWQGMNLERIVALKPDLVIARWGNAEROVDOLASLGI
KVMVDATSEIOIANAROLAPSPQDPKRAEQASLLDOYAKLKAQYADKPKRVFL
QGINPFTSGRESIQNVLEVCGGENIFKDSRPVWPVSREQVLARSQAIYITGSP
DQPKIKQWGEQLKPIVPLTSDMFERASPRILAAQQLCNALSQVD"
complement(9530..10228)
/feature="similar to purine nucleoside phosphorylase (dead)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAB08589.1"
/db_xref="GI:1552737"
/translation="MKIGIGAMEEEVTLRLDKIENROTISLGCGEITYTGOLNGTEVA
LLKSGIGKVAALGATLLEHCKKPDVINTSGAGGLAPTLKVGDIYVSEARYHDADV
TAFGEYQOLPGCPAGFKADDKLIAAAEACIAELNLNAVRLIVSGDATINGSVLAK
IRHNPQAIJAVEMEATAIAHVCHNFVPVVVRAISDVADQQSHLSFDEFLAAKQS
SLWVSLVQLAHG"
10312..11829
/feature="dgt"
/feature="dgt"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="deoxyguanosine triphosphate"
/protein_id="AAB08590.1"
/db_xref="GI:1552738"
/translation="NAQIDFRKKINWHREYRSPQGVKTEHILRIFESDRGRINSIPA
IRLQOQTVFPLERNAAVRTRLTMSMEVQQVGRYIAKEILSRLEKELLEYALGDEL
TGPFESIVEMSCMHDPNPFHGFAGAAINDFORLHPDEASQPLDPRKCSVAAL
RLRDEEPLNELRKIRQDLCEFGNAQGIIRLVHTLMRMLTWAQVGLTKYTRPAW
RGETPETHYLMKPYGILSEAYIARLKRKELNLYSRPLTWIMEAADDISYCVAD
LEDAVEKRIFTVEQLVHHLHEAGQHEKGSLSFLVVENAWEKSRNSLSRSTEDQFFM"

```

```

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x ECU70214 ..
Align seg 1/1 to: ECU70214 from: 1 to: 123171
648 ProPheTyrGluAsnPheTyrAlaGlyGly 657
|||||
30941 CCGTCTACGAGAACTCTATGCGGTGGT 30970

```


seq_name: gb_in:AC009366

seq_documentation_block:

LOCUS AC009366 169839 bp DNA linear INV 05-FEB-2002
 DEFINITION Drosophila melanogaster 3L BAC RP98-48E10 (Roswell Park Cancer
 Institute Drosophila BAC Library) complete sequence.
 ACCESSION AC009366 AC010559 AC010717
 VERSION AC009366.7 GI:18496988
 KEYWORDS HTG:
 SOURCE fruit fly.

ORGANISM

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 169839)

Muzny, D., Scherer, S., Adams, M.D., Holt, R.A., Evans, C.A.,
 Gocayne, J.D., Tabor, P., Williamson, A., Homs, F.H.,
 Dugan-Rocha, S.D., Sodergren, E.S., Hodgson, A.H., Chen, R.C.,
 Ayele, M., Scott, G.S., Worley, K.W., Anammatides, P.G., Brandon, R.C.,
 Rogers, Y., An, H., Baldwin, D., Beeson, K.Y., Brown, M., Buhay, C.,
 Busam, D.A., Center, A., Chen, G., Chen, Z., Clerc-Blankenburg, K.,
 Davenport, L.B., Dietz, S.M., Ding, Y., Dodson, K., Doup, L.E.,
 Draper, H., Emery-Cohen, A., Ferreira, S., Garg, N.D.S., Houck, J.,
 Hostin, D., Howland, T.J., Hume, J., Ibegwam, C., Jalali, M., Kovar, C.,
 Liu, W., Mattel, B., McIntosh, T.C., Morgan, M., Moy, M., Murphy, B.,
 Nelson, K.A., Ndassa, Y., Nguyen, N., Perez, L., Pittman, G.S., Puri, V.,
 Scheeler, F., Shen, H., Strong, R., Tector, C., Wang, Q., Williams, S.M.,
 Wheeler, J., Zaveri, J.S., Zhou, J., Zorrilla, S., Smith, H.O.,
 Xiang, D., Weinstock, G., Gibbs, R. and Venter, J.C.

Unpublished

TITLE

JOURNAL

AUTHORS

2 (bases 1 to 169839)

Worley, K.C., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alsdrooks, S.L., Amaratunge, H.C., Are, J.R., Banks, T., Barbaria, J.,
 Benton, J., Blinag, K., Blankenburg, K., Bonnin, D., Bouck, J.,
 Bowle, S., Brivia, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
 Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
 Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
 Chen, Z., Choudhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
 Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
 Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
 Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
 Dugan-Rocha, S., Durbin, K.J., Earnheart, C., Edgar, D., Edwards, C.C.,
 Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
 Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
 Garza, N., Gill, R., Correll, J.H., Guevara, W., Gunaratne, P., Hale, S.,
 Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
 Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
 Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
 Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
 Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
 Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
 Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
 Loulsegue, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
 Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
 Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
 Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
 Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
 Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S.,
 Ogutu, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
 Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
 Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M.,
 Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,
 Sisson, I., Sodergren, E., Sonake, T., Sparks, A., Stanley, H.,
 Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
 Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
 Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
 Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
 Watlington, S., Williams, G., Williamson, A., Wleczyk, R.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstock, G. and Gibbs, R.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Feb 5, 2002 this sequence version replaced gi:6056147 gi:6056117
 gi:5916439.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the
 entire insert of this clone. Overlapping regions of clones are only
 sequenced and submitted once, so the sequence for the remainder of
 the insert may be found in the record for the adjacent clones.
 Overlapping clones are noted at the beginning and end of the
 Features listing.

ANNOTATION OF FEATURES:

- STSS are identified using ePCR (Genome Res. 7:541-550) searches
 of a local database that includes entries from dbSTS, GDB, and
 local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
 unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
 EST and cDNA sequences. Genes demonstrate at least two exons
 flanked by consensus splice sites that maintained sequence
 continuity across the splice junctions. Sequences that are not
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
 standard of double strand coverage with a minimum of 2 clones and 2
 reads with no ambiguities or 2 chemistries with a minimum of 2
 clones and 3 reads with no ambiguities. If the sequence quality for
 a region does not meet this standard, it will be indicated in the
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
 standards - estimated error rate less than 1 per 10,000 bases.
 Reports of lowest quality individual bases and measures of base
 quality are listed below. Description of the metrics can be found
 at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source

Location/Qualifiers
 1..169839
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="3L"
 /clone="RP98-48E10"
 BASE COUNT 44884 a 39408 c 39261 g 46286 t
 ORIGIN

alignment_scores:

Quality: 10.00 Length: 10
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AC009366 ..

Align seg 1/1 to: AC009366 from: 1 to: 169839

789 SerTyrAlaLysProLeuAsnLysLysGln 798

|||||

99653 TCTATTGCAAGCCACTTACAAAAAGCAA 99682

seq_name: gb_in:AC023676

seq_documentation_block:

LOCUS AC023676 179611 bp DNA linear INV 03-JAN-2002
DEFINITION Drosophila melanogaster 3L BAC RP98-22C11 (Roswell Park Cancer
Institute Drosophila BAC Library) complete sequence.

ACCESSION AC023676

VERSION AC023676.3 GI:17860983

KEYWORDS HTC

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,
Gocayne,J.D., Tabor,P., Williamson,A., Homsí,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M.A., Scott,G.S., Worley,K.W., Amaratunga,H.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Brandon,R.C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Ferreira,S., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwam,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Nodda,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
Schaefer,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstock,G., Gibbs,R. and Venter,J.C.

Direct Submission

Unpublished

2 (bases 1 to 179611)

Worley,K.C.

Direct Submission

Submitted (17-FEB-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 179611)

Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Lucier,A., Luna,R.,
Loulseged,H., Lozardo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,G., Oragunye,N., Nickerson,E., Nwokenkwo,S.,
Ogulu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,H.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N.,
Stinson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Woodson,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Woodson,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovac, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseghe, H., Lozado, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, M., Mawhiney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, S., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogutu, H., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.

Direct Submission
Submitted (03-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Dec 15, 2001 this sequence version replaced gi:6996769.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

TITLE
JOURNAL

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base annotation as Low Coverage.

alignment_scores:

Quality:	10.00	Length:	10
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-701-711-2 x AC023676 ..

Align seg 1/1 to: AC023676 from: 1 to: 179611

789 SerTyraLaLyProLeuAsnLysLysGln 798

|||||

53183 TCTTATGCAAGCCACTTAACAAAAGCAA 53212

seq_name: gb_ba:AJ414154

seq_documentation_block:

LOCUS AJ414154

DEFINITION Yersinia pestis strain CO92 complete genome; segment 14/20.

ACCESSION AJ414154 AL590842

VERSION AJ414154.1 GI:15980810

KEYWORDS

SOURCE

ORGANISM

Yersinia pestis.

Yersinia pestis

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Yersinia.

REFERENCE

AUTHORS

1 (bases 1 to 204050)

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G.,

Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L.,

Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdono-Tarraga, A.M.,

Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G.,

Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S.,

Karlyshev, A.V., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K.,

Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B.G.

Genome sequence of Yersinia pestis, the causative agent of plague

Nature 413 (6855), 523-527 (2001)

2 (bases 1 to 204050)

Parkhill, J.

Direct Submission

Submitted (04-OCT-2001) Submitted on behalf of the Yersinia

sequencing team, Sanger Centre, Wellcome Trust Genome Campus,

Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

Notes:

Details of Y. pestis sequencing at the Sanger Centre are available

on the World Wide Web

(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).

FEATURES

source

1. 204050

/organism="Yersinia pestis"

/strain="CO92"

/db_xref="taxon:632"

/note="biovar: Orientalis"

complement(247..1095)

/gene="YPO2841"

complement(247..1095)

/note="Similar to Agrobacterium tumefaciens mannopine

permease MotD TR:Q44381 (EMBL:AF242881) (273 aa) fasta

scores: E(): 0, 41.5% id in 258 aa, and to Pseudomonas

aeruginosa probable ABC transporter permease PA0606

TR:AAG03995 (EMBL:AF004497) (276 aa) fasta scores: E(): 0,

65.4% id in 260 aa"

/codon_start=1

/transl_table=11

/product="putative ABC transporter permease"

/protein_id="CAC92093.1"

/db_xref="GI:15980811"

/translation="MPKPQALKQKQAWLLRGKWKVFNFSAAILLFLIVPLVIVPL

SNAGSPLSYPMAGFSURVQTFPNSDWLGNLSLLIAPLATLLATGLGLVLAIGL

VGEFGKALVMAVLISPMVAVPVIIAVGFFFEARLSNLSYGLVLAHALIGVPEV

VITVTAVLKSVDQNLRAAASLGSQWRVFKVTLPLIAPGVFSGALFAFVAFSDEVV

VTFLASPRQRTLPKRMFAGIRENLDPTIAAASLMVGASLLLLIIVWELLRRSRRL

HPVSP"

complement(292..357)

/gene="YPO2841"

/note="One of 6 probable transmembrane helices predicted

for YPO2841 by TMHMM2.0"

complement(376..597)

/gene="YPO2841"

/note="Pfam match to entry PF00528 BPD_transp,

Binding-protein-dependent transport systems inner membrane

component, score 59.50, E-value 7.3e-14"

complement(454..519)

/gene="YPO2841"

/note="One of 6 probable transmembrane helices predicted

```

misc_feature
for YPO2841 by TMHMM2.0"
complement(508. .594)
/gene="YPO2841"
/Note="PS00402 Binding-protein-dependent transport systems
inner membrane comp. sign."
complement(583. .648)
/gene="YPO2841"
/Note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(694. .759)
/gene="YPO2841"
/Note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(787. .852)
/gene="YPO2841"
/Note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(967. .1032)
/gene="YPO2841"
/Note="one of 6 probable transmembrane helices predicted
for YPO2841 by TMHMM2.0"
complement(1157. .2422)
/gene="YPO2842"
/Note="Similarity to the N-terminus of Agrobacterium
tumefaciens mannopine permease MotC TR:Q44382
(EMBL:AF242881) (294 aa) fasta scores: E(): 2.7e-29, 41.3%
id in 213 aa, and to the full length Pseudomonas
aeruginosa probable ABC transporter permease PA0605
TR:AAG03994 (EMBL:AE004497) (415 aa) fasta scores: E(): 0,
54.7% id in 397 aa"
/codon_start=1
/transl_table=11
/product="putative ABC transporter permease"
/protein_id="CAC92094.1"
/db_xref="GI:15980812"
/translation="MSQTEMTKPVADERQSGPILRQLRKVKAAAYRRSLILITPL
LLFIVSFIFATILGKSIDNPELRVAMPVTVAVLANPGDALPDEAMFALITDLR
NARNGMLSTAKMSYEDNRYSLSIMSTLRKVPAGPVREALISALPLWGLSTWQS
LQRTQPTTRYLLLSAFDRKVDVSGOIKALPAQOALYVNLRLTLNMAAVYLLICVT
LGYPLAYWLAKQPTGRANLLMLVLLPFTWTSILVKTASWIVLLSGSLNLSLGSRL
TEQPLALVFNAGVYLSMTHILLPLFVLPLVAVMKGISPNVRAAISLGAHPFFAAR
VYVPTQVNGVAGALLVFMMAIGYIITPALLGGPDQMVSYVFAFTTNTNMGMAAA
LGSQLLIIVLLVYVYIRVTRTQAEATH"
complement(1196. .1261)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1277. .1498)
/gene="YPO2842"
/Note="Pfam match to entry PF00528 BPD_transp,
Binding-protein-dependent transport systems inner membrane
component, score 40.70, E-value 3.2e-08"
complement(1328. .1393)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1409. .1495)
/gene="YPO2842"
/Note="PS00402 Binding-protein-dependent transport systems
inner membrane comp. sign."
complement(1493. .1558)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1652. .1717)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(1739. .1804)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted

```

```

misc_feature
for YPO2842 by TMHMM2.0"
complement(2246. .2311)
/gene="YPO2842"
/Note="one of 6 probable transmembrane helices predicted
for YPO2842 by TMHMM2.0"
complement(2469. .3497)
/gene="YPO2843"
complement(2469. .3497)
/gene="YPO2843"
/Note="similar to Agrobacterium radiobacter agropine
permease protein AgtB TR:Q9R701 (EMBL:AF242881) (364 aa)
fasta scores: E(): 9.9e-26, 30.3% id in 320 aa, and to
Pseudomonas aeruginosa hypothetical protein PA0222
TR:AAG03611 (EMBL:AE004460) (352 aa) fasta scores: E(): 0,
56.1% id in 342 aa"
/codon_start=1
/transl_table=11
/product="putative transport system periplasmic protein"
/protein_id="CAC92095.1"
/db_xref="GI:15980813"
/translation="MFKIMTLTLLAGWVFOHAEFTLVVSFGGLNKDAQVKAFVKPF
MAAGKTIEAGEYNGEMARIRAMVETGQIGWDIVVEGPELLRGNEGLFIDWSKL
GDERQFIRGTVSQCGSLFSLMMLLIYDASRKTPTVSWADFNWVTTTFGRSLRKRTA
KFTMEIALLADGVKREDYRVLTATPEGVRFVFNKLDQKPNQWWSGAQPLQWLVSF
DVMGVSYNGRVGSALKEGHDFRMYWTDTSIYDMDSWTIVKSKKVLAPQFLAFANLA
ENKVFENIAYGPTNINATLMIDPEIANLPTAPANLAEPFMMNQFVHSEDLQ
RFSNWAAR"
complement(3550. .4839)
/gene="YPO2844"
/Note="gabt; YPO2844"
complement(3550. .4839)
/gene="YPO2844"
/EC_number="2.6.1.19"
/Note="similar to Escherichia coli 4-aminobutyrate
aminotransferase GaoG SW:GOAG_ECOLI (F50457) (421 aa)
fasta scores: E(): 0, 65.7% id in 428 aa, and to
Pseudomonas aeruginosa 4-aminobutyrate aminotransferase
Gabt TR:AAG03655 (EMBL:AE004465) (426 aa) fasta scores:
E(): 0, 53.0% id in 423 aa"
/codon_start=1
/transl_table=11
/product="putative aminobutyrate aminotransferase"
/protein_id="CAC92096.1"
/db_xref="GI:15980814"
/translation="MSHRELQRRKEATPRGIGVLCDYFAVRAENATLWDEQREYID
FTAGIATNIGHRHHPKVMAAVRQQLDQFTHTAYQVVPVAVYVTLAEKINSAPISDSN
YKAGNSKTAFTETGVEAIEANVKIARAATGPGPVAFSGAFHGTLLAMALTGRAVP
YKVGPPAPASIFPHALYNELYGVSVEAISVERLFRCDISPTQVAAILEPPIQEGG
GFNIAPPEVSALRFLCDEHGILLIADDEVQTFARTGKLFAMEYYPDTKVDVITWAKS
LGGMPISAVTGRADIMDALPLPSLGGTYAGNPLAVALDIIAEKLCERALLIL
GAKLVLEKQAQMSNAIIVGRAGSVAVFNDPVSQKPSPELTPAYQRALEBGLL
LLSCGVHSNVRFLYPLTIPDKQKQAMNILLRLAS"
complement(3613. .4749)
/gene="YPO2844"
/Note="Pfam match to entry PF00202 aminotran_3,
Aminotransferases class-III pyridoxal-phosphate, score
459.70, E-value 2.4e-159"
complement(4000. .4116)

```

```

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AJ414154
..
Align seg 1/1 to: AJ414154 from: 1 to: 204050
599  GlnGlyMetSerHisSerValAspLeuThr 608
|||||
90253 CAAGGAATGAGTCACCTCTGTGAGACCTGACT 90282

```

```
seq_name: gb_ba:AJ414146
seq_documentation_block:
LOCUS AJ414146 210050 bp DNA linear BCT 04-OCT-2001
DEFINITION Yersinia pestis strain C092 complete genome; segment 6/20.
ACCESSION AJ414146 AL590842
VERSION AJ414146.1 GI:15979072
KEYWORDS
SOURCE
ORGANISM
Yersinia pestis.
Yersinia pestis
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Yersinia.
1 (bases 1 to 210050)
Parkhill,J., Wren,B.W., Thomson,N.R., Titball,R.W., Holden,M.T.G.,
Prentice,M.B., Sebahia,M., James,K.D., Churcher,C., Mungall,K.L.,
Baker,S., Basham,D., Bentley,S.D., Brooks,K., Cerdeno-Tarraga,A.M.,
Chillingworth,T., Cronin,A., Davies,R.M., Davis,P., Dougan,G.,
Feltwell,T., Hamlin,N., Holroyd,S., Jagels,K., Leather,S.,
Karlshede,A.V., Moule,S., Oyston,P.C.F., Quail,M., Rutherford,K.,
Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrall,B.G.
Genome sequence of Yersinia pestis, the causative agent of plague
Nature 413 (6855), 523-527 (2001)
21470413
2 (bases 1 to 210050)
Parkhill,J.
Direct Submission
Submitted (04-OCT-2001) Submitted on behalf of the Yersinia
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
Notes:
Details of Y. pestis sequencing at the Sanger Centre are available
on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/Y_pestis/).
FEATURES
source
1. 210050
/organism="Yersinia pestis"
/strain="C092"
/db_xref="taxon:632"
/notes="biovar: Orientalis"
829..11946
/genes="yapH"
/genes="yapH"
/notes="Previously sequenced as Yersinia pestis YapH
protein TR:CAC14227 (EMBL:AJ277631) (3705 aa)
scores: E(): 0, 99.6% id in 3705 aa. Similar in regions to
Escherichia coli Ycha protein TR:09JMS3 (EMBL:AP001918)
(1371 aa) fasta scores: E(): 0, 39.1% id in 1405 aa, and
to Escherichia coli adhesin Aida-I precursor SW:AIDA.ECOLI
(Q03155) (1286 aa) fasta scores: E(): 0, 29.2% id in 1286
aa. Contains a possible N-terminal signal sequence."
/codon_start=1
/transl_table=11
/product="putative autotransporter protein"
/protein_id="CAC89847.1"
/db_xref="GI:15979073"
/translation="MTIFKIVNNSINVVVSELAKEGRITKSRNLISEGVLPKF
EQSWSKLFRKLLSLGSLVFLSGPFAADITVSTQELSAALNSGYIKIILGA
DITLGLSVNMTSNQVLDGQFGLTVNNTTNYGLVSSGSLTTLQNMKSIDNSA
YYSMVNLGANTAVNYIYNNIDFLGSLIYMGAYATNSIMTFGDLNDVVVNDRA
QETGVNKAFTAFADGAFELIANQVFSGTTRNGLEIGSYNSIDFGSGQVILQSR
TSPGVNAFTAFADGAFELIANQVFSGTTRNGLEIGSYNSIDFGSGQVILQSR
SDGSIISGNDIATNNAAGINNNSAGDANVYINLGTGSLIKATNTGILATKNNANS
DIYIRAGDTAIGATGATHNGTGIKINDGTTITTAGIASSAISISVDNTDG
TITPAGTGVNVLASAILNLFPGTINTSATNGITFAGTGGHTLDTLILNLTGIA
LSNVAQVNLTSVNTLNLGTLNLSLTGLTVDSLNGRNTINIEGAGIGATNTEL
NTPDACALDINVGAGTIGQATGGVNLGSLNLIINVATLQITDIDNTTIG
NETQLAENATAINEIGSSSKTLNNNGTIKGSVIFAGVADHIINNGTLDGTLTGAG
NLDAGLGLTNDVNLGDNNSVTIQGATVNSVLIITGNNGDFTFINGMSVSTTIG
SLDAGLGLTNDVNLGDNNSVTIQGATVNSVLIITGNNGDFTFINGMSVSTTIG
ELFGSTFGIHLATGAGTSAIVNNSASLEQSMFAGTQVQVNOGALTSASNQ
LGSAKIGLGLTNDVNLGDNNSVTIQGATVNSVLIITGNNGDFTFINGMSVSTTIG
```

```
DLLSAETRVKTAENQDFSRWLSQSPWKSVVORIIEPRYAAAVAKOYHALENIYDPKL
SP"AAELAAANGTGDVDANRIVGRKINDELGMELDMALTHEVLSAKGASSLLDNLWMEYLI
misc_feature
/gene="yPO1005"
/length=13164
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 9.30, E-value 44"
misc_feature
/complement(13219..13290)
/gene="yPO1005"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 12.80, E-value 8.4"
misc_feature
/complement(13294..13353)
/gene="yPO1005"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 10.40, E-value 30"
misc_feature
/complement(13468..13539)
/gene="yPO1005"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 17.30, E-value 0.36"
misc_feature
/complement(13543..13602)
/gene="yPO1005"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 12.90, E-value 7.8"
gene
/complement(14492..15367)
/gene="yPO1006"
CDS
/complement(14492..15367)
/gene="yPO1006"
/notes="Similar to neighbouring CDSs YPO1005 and YPO1007.
Similar to regions of Salmonella typhimurium secreted
protein h2 TR:O9RPH0 (EMBL:AF160727) (788 aa) fasta
scores: E(): 2.5e-24, 41.9% id in 272 aa, and to regions
of Yersinia pestis outer membrane protein YopM TR:068701
(EMBL:AF053946) (409 aa) fasta scores: E(): 7e-23, 40.5%
id in 259 aa"
/transl_table=11
/product="putative antigenic leucine-rich repeat protein"
/protein_id="CAC89849.1"
/db_xref="GI:15979075"
/translation="MYLSNITSNVSNMIPGDRRETHADRPAAATALTADYHAIWEKWE
NDPRVAGEQVAGVARKKECLNNAERLNLSLDTLSLPTLPCNELNICNLTE
LPTLPDLNLTAKSYNOLRTLPNTLPASLLSLKVMNELERLPELPEGLKLDVGC
NTSLQPLPVLPSLELISNCLTEPLTLPNSLKELDAHGQLRLDPLPISLLRL
NVAYNQLALPENLPGSLRCIYTEYNQLSQLPDLAHLRQNCNCLDGNPLSPFLLAL
LRLSTKPNYQGPRI"
/complement(14657..14728)
/misc_feature
/gene="yPO1006"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 15.40, E-value 1.3"
misc_feature
/complement(14732..14791)
/gene="yPO1006"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 9.30, E-value 43"
misc_feature
/complement(14981..15040)
/gene="yPO1006"
/notes="Pfam match to entry PF00560 LRR, Leucine Rich
Repeat, score 11.30, E-value 22"
alignment_scores:
Quality: 10.00 Length: 10
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x AJ414146 ..
Align seq 1/1 to: AJ414146 from: 1 to: 210050
648 ProPheTyrGluAsnPheTyrAlaGlyGly 657
|||||
76307 CCATCTACGAAAACCTCTATGACGGTGT 76336
seq_name: gb_ba:AL627266
```

```
seq_documentation_block:
LOCUS AL627266 268050 bp DNA linear BCT 25-OCT-2001
DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
complete chromosome; segment 2/20.
ACCESSION AL627266 AL513382
VERSION AL627266.1 GI:16501496
KEYWORDS
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 268050)
AUTHORS
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Church,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Conerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrar,J., Felwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
Stevens,K., Whitehead,S. and Barrell,B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18
NATURE 413 (6858), 848-852 (2001)
PUBMED 11677608
REFERENCE
2 (bases 1 to 268050)
AUTHORS
Parkhill,J.
TITLE
Direct Submission
JOURNAL
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK
COMMENT
E-mail: parkhill@sanger.ac.uk
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
( URL, http://www.sanger.ac.uk/Projects/S_typhi/ ).
FEATURES
source
1. .268050
Location/Qualifiers
/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/strain="CT18"
/db_xref="taxon:90370"
181. .906
/gene="STY0239"
/notes="rpsB"
181. .906
/gene="STY0239"
/notes="Orthologue of E. coli rpsB (RS2_ECOLI); Pasta hit
to RS2_ECOLI (240 aa), 98% identity in 240 aa overlap"
/codon_start=1
/transl_table=11
/product="30S ribosomal protein S2"
/protein_id="CAD08674.1"
/db_xref="GI:16501497"
/translation="MATVSMRDLKAGVHFQHTRYWNPMPFFIGARKKVIINLE
KTVPMFEALAEINLKISARKKILFVGTAKRAAEAKAANSCDQFFVNRHWGGMLT
NWKTVQSIKRLKIDLETQSDGTFEKLTKEALMRTELEKLSIGIKDGMGLPDA
LFVIDADHEHIAIKDANNILGIPVFAIVDTNSDPDGVDFVPGNDDAIRAVSLYLGA
VATVREGRSQDLASQAESEFVEAE"
196. .231
/misc_feature
/gene="STY0239"
/notes="STY0239"
205. .855
/misc_feature
/notes="PS00962 Ribosomal protein S2 signature 1"
/gene="STY0239"
/notes="Pfam match to entry PF00318 Ribosomal_S2, Ribosomal
protein S2, score 441.70, E-value 6.5e-129"
552. .726
/misc_feature
/gene="STY0239"
/notes="PS00963 Ribosomal protein S2 signature 2"
1164. .2015
/gene="STY0240"
/notes="tsf"
1164. .2015
CDS
```

```

/genes="STY0240"
/note="Orthologue of E. coli tsf (EFTS_ECOLI); Fasta hit
to EFTS_ECOLI (282 aa), .97% identity in 282 aa overlap"
/codon_start=1
/transl_table=11
/product="elongation factor Ts"
/protein_id="CAD08675.1"
/db_xref="GI:16501498"
/translation="MAEITVASLVKELRERTGAGMDCKKALTEANGDIELATENMRKS
GAIKAAKAGNVAADGVITKIDGNVAFILEVNCQDTEVAKDAGQAFADKVLDRAVA
GKITDEVLRADGFEERVALVAKENINIRRVASLEGDLVGSYQHGARGVILRAAKG
ADEELVKQALMAHAKSEFVPEDVSADVKEQVOLDIAMQSKPKKEIAEKVVEG
RNMKFTGVSLSGQCFVMEPSKSVGLLKEHNADVTFIRFEVGEIGEKVETDFAEV
AAMSQS"
misc_feature
1170..1289
/genes="STY0240"
/note="Pfam match to entry PF02094 TS-N, TS-N domain,
score 81.00, E-value 2.4e-20"
misc_feature
1197..1244
/genes="STY0240"
/note="PS01126 Elongation factor Ts signature 1"
misc_feature
1332..1952
/genes="STY0240"
/note="Pfam match to entry PF00889 EF-TS, Elongation
factor TS, score 356.50, E-value 2.9e-103"
misc_feature
1386..1418
/genes="SPY0240"
/note="PS01127 Elongation factor Ts signature 2"
gene
2160..2885
/genes="STY0241"
/note="pyrH"
CDS
2160..2885
/genes="STY0241"
/note="Orthologue of E. coli pyrH (PYRH_ECOLI); Fasta hit
to PYRH_ECOLI (240 aa), 98% identity in 240 aa overlap"
/codon_start=1
/transl_table=11
/product="uridine 5'-monophosphate kinase"
/protein_id="CAD08676.1"
/db_xref="GI:16501499"
/translation="WATNAKPVYKRIKLLKSGEALQGTGEGIDASILRMAQEIKEI
VELGTQGVVIGGNLFRGAKLAKAGMNVVGDHGMGLATVNGMLMRDALHRAVYNA
RLMSAIPNGVCDNTSWAEASILLRNRRVYVILSAGTGNPFTTDSACLRGIEIEADV
VLKATKVDGVFTADPAKDPASATMYDQLTYSVLDKELKVMDLAAFTLARDHKLPVRF
NNKPCALRRVYVNGEKGSLITE"
misc_feature
2187..2738
/genes="STY0241"
/note="Pfam match to entry PF00696 aakinasase, Amino acid
kinase family, score 153.90, E-value 2.9e-42"
gene
3032..3389
/genes="STY0242"
/note="frr"
CDS
3032..3389
/genes="STY0242"
/note="Orthologue of E. coli frr (RFRF_ECOLI); Fasta hit to
RFRF_ECOLI (185 aa), 94% identity in 185 aa overlap"
/codon_start=1
/transl_table=11
/product="ribosome recycling factor"
/protein_id="CAD08677.1"
/db_xref="GI:16501500"
/translation="MTSDIRKDAEVRMEKVEAFKTOISKVRTGRASPSLLDGIVVEY
YGTPTPLROLASTVTEDSRTKLINVDKSMGPAVEKAIINASDLGLNPPSAGDVRVPL
PLTEERRDLKIVRGEARQAVRVNRDANDKVKALLDKDKAISSEDDRRSQEEV
QKMDNAIKKVDAAADKAEELMQF"
misc_feature
3086..3380
/genes="STY0242"
/note="Pfam match to entry PF01765 RRF, Ribosome recycling
factor, score 339.90, E-value 2.9e-98"
gene
3731..4927
/genes="dxr"
/note="STY0243; yaem"
CDS
3731..4927
/genes="dxr"
/EC_number="1.1.1.1"
/note="Similar to Escherichia coli 1-deoxy-D-xylulose
5-phosphate reductoisomerase dxr SW:DXR_ECOLI (P45568;
P77209) (398 aa) fasta scores: E): 0, 88.9% id in 397 aa
Orthologue of E. coli yaem (DXR_ECOLI); Fasta hit to
DXR_ECOLI (398 aa), 89% identity in 397 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein in frr 3'region"
/protein_id="CAD08678.1"
/db_xref="GI:16501501"
/translation="MKQLTILSTGSGICSTLDVVHHPDSFRVIALVAGKNVARMAE
QCLEFSPYAVMDTTSAEOLKIMLQOHSRTEVLGQQAACMAALDEVGHVMAAIV
GAGLLPTLAAIRAGKTILLANKESILVTCGRLEFMDVKRSNARLLPVDSEHNAIFQSL
PQSIQHNLYADLEQNGVTSILLTSGGGPRETPMCDLAAMTPOACRHPNWSMGKRI
SVDSATMMNKGLEYIEARWLFNASRQMEVLIHPQSVIHSNRYQDSVLAQLGEPDM
RTPIAHTMAMPNRYTGAQPLDFCKLSALTFSAPDYQRYPCCLKAMEAFEGQQAATTA
LNAANEITVAEFLAQQIRFTDIAGNLAVLERMDLHEPASVDDVLQYDAIAREVARKO
VIRLSR"
misc_feature
4064..4087
/genes="dxr"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
gene
5240..5998
/genes="STY0244"
CDS
5240..5998
/genes="STY0244"
/note="Orthologue of E. coli UPPS_ECOLI; Fasta hit to
UPPS_ECOLI (253 aa), 94% identity in 252 aa overlap"
/codon_start=1
/transl_table=11
/product="undecaprenyl pyrophosphate synthetase"
/protein_id="CAD08679.1"
/db_xref="GI:16501502"
/translation="MLSATQPVSENLPAHGCRHVAIIMDNGRWAKQKQIRAPGHKA
GAKSVRAVSFAANGIDALTLYAFSENWNRPAQEVSAIMELFWALDSEVKSLLHRL
NVRLRIGDISREFNSRLQERIRKSEALTHTGLTLNIAANYGGWMDIVQGVRLAEL
VOAGVLRPQDIDERLQOQICMHELAPVDLIVIRTGGEHRISNFIWMQIAYAELYFTDV
LWPDFDEQFEGALHAFANRERFEGTGPCDDKA"
misc_feature
5306..5971
/genes="STY0244"
/note="Pfam match to entry PF01255 UPP_synthetase,
Putative undecaprenyl diphosphate synthase, score 492.70,
E-value 2.8e-144"
misc_feature
5804..5857
/genes="STY0244"
/note="PS01066 Undecaprenyl pyrophosphate synthetase
family signature"
gene
6011..6868
/genes="STY0245"
/note="cdsA"
CDS
6011..6868
/genes="STY0245"
/note="Fasta hit to YNBB_ECOLI (298 aa), 33% identity in
298 aa overlap
Orthologue of E. coli cdsA (CDSA_ECOLI); Fasta hit to
```

```

alignment_scores:
  Quality: 10.00      Length: 10
  Ratio: 1.000       Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-701-711-2 x AL627266
  Align seg 1/1 to: AL627266 from: 1 to: 268050
  648 ProPheTyRGluAsnPhetYrAlaGlyGly 657
  |||||
  10187 CCGTTCATGAGACTTCTATGCCCGTGGT 10216

seq_name: gb_ba:AP002550
```

```
seq_documentation_block:
LOCUS      281530 bp      DNA      linear      BCT 07-MAR-2001
DEFINITION Escherichia coli O157:H7 DNA, complete genome, section 1/20.
ACCESSION AP002550 BA000007
VERSION   AP002550.1 GI:13359456
KEYWORDS
SOURCE     Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RMD 0509952)
           DNA.
ORGANISM   Escherichia coli O157:H7
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE  1 (sites)
AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
           Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,T., Abe,H., Iida,T.,
           Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
           Sasaki,C. and Shinagawa,H.
TITLE     Complete nucleotide sequence of the prophage Vm2-Sakai carrying the
           verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
           derived from the Sakai outbreak
JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE   20198780
REFERENCE  2 (sites)
AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
           Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
           Hayashi,T.
TITLE     Comparative analysis of the whole set of rRNA operons between an
           enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
           Escherichia coli K-12 strain MG1655
JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE   20557356
REFERENCE  3 (sites)
AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
           Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
           Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasaki,C. and
           Shinagawa,H.
TITLE     Complete nucleotide sequence of the prophage Vm1-Sakai carrying the
           Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
           O157:H7 strain derived from the Sakai outbreak
JOURNAL   Gene 258 (1-2), 127-139 (2000)
MEDLINE   20564182
REFERENCE  4 (sites)
AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
           Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
           Tanaka,M., Toke,T., Iida,T., Takami,H., Honda,T., Sasaki,C.,
           Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
           Shinagawa,H.
TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
           O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL   DNA Res. 8 (1), 11-22 (2001)
MEDLINE   21156231
REFERENCE  5 (bases 1 to 281530)
AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
           Hayashi,T.
TITLE     Direct Submission
JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
           Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
           Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
           URL:http://www.gen-info.osaka-u.ac.jp/,
           Fax:81-6-6879-2047)
COMMENT   genome project.
FEATURES   Location/Qualifiers
           1..281530
           /organism="Escherichia coli O157:H7"
           /strain="O157:H7"
           /sub_strain="RMD 0509952"
           /db_xref="taxon:83334"
           190..273
           /gene="Ecs0001"
           190..273
           /gene="Ecs0001"
           /note="similar to THRL_ECOLI gi11786182 percent similarity
           100 in 21 aa, but has 6 additional residues (Conserved in
           E.coli K-12)"
gene
CDS
```

```
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="thr operon leader peptide"
/protein_id="BAB33424.1"
/db_xref="GI:13359457"
/translation="MKRISTTTTITTTTITTTITTTITTTGAG"
354..2816
/gene="Ecs0002"
354..2816
/gene="Ecs0002"
/note="similar to THRA_ECOLI gi11786183 percent identity
99 in 820 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="aspartokinase I-homoserine dehydrogenase I"
/protein_id="BAB33425.1"
/db_xref="GI:13359458"
/translation="MRVLKFGTGSVANAERFLRVADILESNAEQGVATVLSAPAKIT
NHLVAMIEKTIISGDALPNISDAERIFAEALFTGLAAQGFPLAQIKTVEDEFFAQIK
HVLGHSILGQCPSINAALICRGEKMSIAIMAGLEARGHNVTIPDVEKLLAVGHY
LESTVDIAESTRRIAASRIPADHVMVMAGFTAGNEKGLVILGRNGSDYSAVLAACL
RADCEIWTDDVGVTCDPQVPDARLLKMSYQEMELSYFGAKVLPHTTPTIAQF
QIPCLIKNTGNPAFGTLLIGASRDEDELPKGISLNNNMFVSQPCMKGMVMAAR
VFAAMSRARISVVLITQSSEYSISFCVPQSDCVRAERAMQEEFYLELKEGLEPLAV
TERLAIISVVGDMRTLRGISAKFFAALARANINIVIAOGSSERSISVYVNNDDATT
GVYRTHOMLFNTDQVIEVIGVGGVGGALLBQLKQQSWLNKKHIDLRVCGVANSKA
LLTNVHGLNLENQOEELAQAPFNRLIRLVKHLNLPVIVDCTSSQAVADYVAD
FLREGFHVTPNKKANTSMDYHLLRHAAEKSRKFLYDTNVGAGLPVNIENLNLN
AGDELMKFSGLISLSIFGKLDGMSFSEATTIAREMGYTEPPDRDLSDMDVARK
LLILARETGRELELADIEIEPVPFAEFNAEGDVAAPMANLSQDDLLFLARVAKARAK
KVLRYGVNIDEGVCVRIAEVNDGNDPLFKVKNGENALAFYSHYYOPLPLVLRGYGAG
NDVTAAGVFADLLRTLTSWKLGV"
2818..3750
/gene="Ecs0003"
2818..3750
/gene="Ecs0003"
/note="similar to THRE_ECOLI gi11786184 percent identity
99 in 310 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="homoserine kinase"
/protein_id="BAB33426.1"
/db_xref="GI:13359459"
/translation="MVKYVAPASSAMNSVGFVDLGAATVPDGLIGDVVTVESAEFTF
SLNNLGRFADKLPSREPNIVYQCWERFCELGKQIPVAMTLKKNPIGSGLGSSACS
VVALMANNEHCCKPINDTRLALMGELEGRISGSHYDNVAPCTGLGGMLTEENDI
ISQVPQFDEWLVLAYPGIKVSTAERAILPAQYRRQDCIAHGRHLAGFIHACYSRQ
PELAAKIMKDVIAEPIREKLPGFRQARQAAVEIGAIVASGISGSGPTLFCALCKPDTA
QRVADWLKNTYLQNGEGFVHICRLDTAGARVLEN"
3751..5037
/gene="Ecs0004"
3751..5037
/gene="Ecs0004"
/note="similar to THRC_ECOLI gi11786185 percent identity
99 in 428 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/product="threonine synthase"
/protein_id="BAB33427.1"
/db_xref="GI:13359460"
/translation="MKLYNLKDNHNEQVSFAQAVTOGLGKNQGLTFPHDLPEFSLTEID
EMLKIDFVTRSAKILSAFIGDEIPQEIILFEERVAAFAFAPVAVNVEDYVGCLELPHGP
TLAFKDFGGRFMAQMLTHIAGDKPVTILLTSGDTGAAYAHAFYGLPNKVYILYPRG
KISPLQEKLFCTLGNIETVAIDGDFDACQALVQKQAFDDEELKALGLNSANSINLSR
LLAQICYFEAVQAQLPQARNQLVSVSPNGFDLTAGLSKLGLPLVKRFRITAAATVNW
DTPVRFHLDGQSPKATQATLSNMDVSNQNNPVEELFRKIKWLKELGAAVADDE
TTQQPMRELKELGYTSEPHAAVAYRALRDLQNPGEYGLFLGTAPAKFESVEAILGE
TLDLPKELAEARADLPLLSHNLPAFALRKLMMNHQ"
5251..5547
gene
```


CDS	/gene="ECs0005" 5251..5547 /gene="ECs0005" /note="similar to B0005_ECOLI gi 1786186 percent identity 95 in 98 aa (Conserved in E.coli K-12)" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB33428.1" /db_xref="GI:13359461" /translation="MKKMSIVLALSILVLVAPMATQAAEITLVPVKLQIGDRNRGY YWDGHRDHGMWKQYEWNRGHPHPPPPPRHHKKAHHDHGGHGGPKGKHHR" complement(5700..6476) /gene="ECs0006" complement(5700..6476) /gene="ECs0006" /note="similar to YAAA_ECOLI gi 1786187 percent identity 99 in 258 aa (Conserved in E.coli K-12)" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="hypothetical protein" /protein_id="BAB33429.1" /db_xref="GI:13359462" /translation="MLILSPAKTLDYQSPLTTRTYTLPELDNSQOLIHEARKLTPP QISTLMRISDKLAGINAAFRHDMQPDFTENARQAAILAFKGDVYTGLOAETFESEDDFD FAQOHLRMISGLGYLRLPDLMPYRLEMGIRLENARGKDLQYFGWGDITNKLNEALA AOGDNVYNLASDEYFKSVKPKKNAEIIKPFLVDENKGRFKIISFYAKKARGLMSRF IENRLTKPQLGFPNSEGFFDEASSNGELVFKRYEQR" complement(6546..7976) /gene="ECs0007" complement(6546..7976) /gene="ECs0007" /note="similar to YAAJ_ECOLI gi 1786188 percent identity 99 in 476 aa (Conserved in E.coli K-12)" /codon_start=1 /transl_table=11 /evidence=not_experimental /product="putative inner membrane transport protein" /protein_id="BAB33430.1" /db_xref="GI:13359463" /translation="MPDFESFINSLVMSVMIYLLFGAGCWFTFRTGFQVFIYRQFG KSLKNSIHPPQGLTSFQSLCTSLAARVSGNLAGVALAITAGPGGAVFWMMVAAFG MATSFACESLAQLYKERDVNGOFRGGPANTMARGLMRWGVLFAVFLIAYGLIFSG VOANAVARALSSEDFPPLVGTGILLAVEALLIIRGLHGVARLMQGFVPLMAIIWLT SLVICVINIGOLPHVWISIFESAFCHQEAAGNAGTYLSQATNGFQRFMSNEAGMG STPNAAAASWPPHPAAGTIVOMIGIFIDTIVICTASAMILLIAGNAGTYMPLEGIQ LLOKARVLMWSGAEFVTLVILFAFSSIVANYIYAENNLFFLRNNPKAIWCLRIC TFATVIGGTLTSLPLMWQLADIIMACMAITNLTAILLSPVVHTIASDYLRRKLGVR PVFDLPYDPDIGRLSRDWDVDSQ"	LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS FEATURES source mRNA gene	AE003597 Drosophila melanogaster genomic scaffold 142000013386036 section 7 of 9, complete sequence. AE003597 AE002647 AE003597.1 GI:7296525 HTG. fruit fly. Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 295566) Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Ananides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, Y.H., Blazer, R.G., Champe, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K.C., Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrieri, S., Fleischmann, W., Fosler, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstock, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C. The genome sequence of Drosophila melanogaster Science 287 (5461), 2185-2195 (2000) 20196006 2 (bases 1 to 295566) Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J. Direct Submission Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA Location/Qualifiers 1. 295566 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /chromosome="3L" join(<27808..27909,28034..28318,28822..>29085) /gene="CG12673" /product="CT35415" /db_xref="FLYBASE:FBan0012673" /db_xref="FLYBASE:FBgn0037153" /evidence=not_experimental <27808..>29085 /gene="CG12673" /map="79D4-79E1" /db_xref="FLYBASE:FBan0012673"
alignment_scores	Quality: 10.00 Length: 10 Ratio: 1.000 Gaps: 0 Percent Similarity: 100.000 Percent Identity: 100.000		
alignment_block	US-09-701-711-2 x AP002550 ..		
Align seg 1/1 to: AP002550 from: 1 to: 281530			
648 ProPheTyrGluAsnPheTyrAlaGlyGly 657 			
203210 CCGTCTCAGAGAACTCTATGCCGGTGGT 203239 			
seq_name: gb_in:AE003597			
seq_documentation_block:			

```
/db_xref="FLYBASE:FBgn0037153"
/feature=not_experimental
join(27808..27909,28034..28318,28822..29085)
/gene="CG12673"
/note="CG12673 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012673"
/db_xref="FLYBASE:FBgn0037153"
/evidence=not_experimental
/protein_id="AAF51810.1"
/db_xref="GI:7296526"
/translation="MGRVGRRAFGPRGAPPAQRCIRVHVSNNDIIAANDTMRLLMY
HAQPPHGVSRVGLPDPAPARFPRMVLQRAQLPMPSTHDERVRVLELRNEDVE
LPADLDFKMFKEIDINKRHLRIEPIYDSSSVHYLIQHTLHECOGAHALEEE
MARQGPICGARGIPACNAIVASWSRGSEVTSIRERNQNLPPRYPCQNSCPT"
join(<35333..35500,35594..35756,35909..>35937)
/gene="CG7495"
/product="CT23023"
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
<35333..>35937
/gene="CG7495"
/map="79E1-79E1"
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
join(35333..35500,35594..35756,35909..35937)
/gene="CG7495"
/note="CG7495 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0007495"
/db_xref="FLYBASE:FBgn0037154"
/protein_id="AAF51811.1"
/db_xref="GI:7296527"
/translation="MDPNWRHIIPPGOKRVVSEQCIEDCTGYAFPOGGINIAVMNR
THQIGKVKLRQIQTELPPIADHSNIDVAYQDFRLPQSVHMPGDRLLAECIYDS
SSRAITIGDEVPAIS"
join(<48427..48474,48682..48989,49294..>49318)
/gene="CG14461"
/product="CT34144"
/db_xref="FLYBASE:FBan0014461"
/db_xref="FLYBASE:FBgn0037155"
/evidence=not_experimental
<48427..>49318
/gene="CG14461"
/map="79E1-79E1"
/db_xref="FLYBASE:FBan0014461"
/db_xref="FLYBASE:FBgn0037155"
/evidence=not_experimental
join(48427..48474,48682..48989,49294..49318)
/gene="CG14461"
/note="CG14461 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0014461"
/db_xref="FLYBASE:FBgn0037155"
/evidence=not_experimental
/protein_id="AAF51812.1"
/db_xref="GI:7296528"
/translation="MWAEISDGTTLCSLSEFCYSINVTKYKKHRRCKPKRPL
APPTERTAPPASDLSELPLVHLDNNNIEGARSSRSATDVHLSRSGSRFISC
LLWLGSWLLMLLRTPPTPTC"
complement(join(64112..66736,66987..69059,69117..69239))
/gene="BcdNA:GH03694"
/product="CT26020"
/db_xref="FLYBASE:FBan0009063"
/db_xref="FLYBASE:FBgn0028500"
complement(<64112..>69239)
/gene="BcdNA:GH03694"
/note="CG9063"
/map="79E1-79E1"
/db_xref="FLYBASE:FBan0009063"
/db_xref="FLYBASE:FBgn0028500"
complement(join(64502..66736,66987..69041))
/gene="BcdNA:GH03694"
```

```
/note="BcdNA:GH03694 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0009063"
/db_xref="FLYBASE:FBgn0028500"
/protein_id="AAF51813.1"
/db_xref="GI:7296529"
/translation="MYTPVGPWKRVGLALPGESASIRIHCCDAVKILVAAYGDDPLGI
WYANLPIAYFRTEDSLQYGANOLIVKRPDSRQALTLASGLLQYLDLFEANGM
GILQIDPPAASLRDAELFKENISRLSLRCLSVTLGSVITVCCISLSEILLAT
QSCELLRLWTELEHAENDLELPALSKIKLRDIPFYVQOQQAQARNVPLNRDSYVA
SLEYSPITGCCAAVFSDDRAFLJANHLEFETHMGCEWMPDVEDASVCSYNHFRLL
AYGESSAVKYAIDDAITGGLFESHRLITENILPDSLSVNEKLKWSPDGCVLAVSWT
NGSLWSTFGALLMSTLWDFNQNDVLCQNPILKIRLEWSTGYLFMLKLHPKED
KSNYLQFQVKSALSNMPCWTTSPHLLQGDCLYLQGNNELETYAGSHGTFPSGL
GSDDEDISGDCLSELKQSPHTGSITLESKYWTVLQPLNYAATNWPRIYAAIDPGLH
LAVAGRTGLAHYSLVTRRWKLFGENSOEKDFVSGGLLWHGVVMGYSLLDRTDEL
RCYPADCKLDNOYGHKLQVRAPIVSLANSFRHOLIVLTADGIVSLFNMSKNASALDIE
CAYELDKVSCIHPACIVSLVTNLKNELKFPQGLGQDAETIIVNCGRILMQORDA
GEOQPNTLALTCLASCEVEFWLSHLERCAMRDCWLVSAGHMRWLPIPPGRERR
GEOGQARLHSMFKRIMLSFPLKPLVVLFDNVIYLVGENESTLYANQVSHFSL
PFAYMERKSOIYLHKVRLQIKRNLGYSANEMASOCSGLPYFPHALELLHEVLEEEA
TSKOPIPDAOLPSILDFIRFPVYLETIVQCARTEIATLWPLFLFSMAGKPKDLQMCCL
QSEQLDTAASVLIILQNLSPSVSKQYATMLLDIALQQRKWEAKDLIRFLKADPNE
IDSPSSMVNVKIAPPVQNTQQVQONDAFNMVLGPIARERSFTVTNSLPKOK
QASGTPGVAPVTESAGAPSVRRSTKQRETCIDILQHRARQLQNHKMLDLGY
MCAYDLFHLWSLQSESERAAKDDFAGALQALHEELDLPFPFAAKDDPAQIRGSL
ROTCGGSSQTSSECYFSLATPNGAATQSPOLQPSIRREELELOQPSLPLKTRSGS
QLSPDNFRYRLYSLPTSEDDLAVDILPQKLSIKRLVLLQLFIBANTDYALVLSILL
QDAASIRVNGIIRSESVHTRCTESALKQLSQSTFEHSGSLYRGFVLTLOPHVYLL
EQYIOSIGDAPCSQLODAGPTGQVDVSTGLQNGEVGFVPSQOANQWTVAD
LNPQHRLTRHASLESNGNAVASGSAHSPTQRLSRONSREGRCLM"
join(69753..69861,69916..69967,70021..71690,71763..>72208)
/gene="Ddx1"
/note="Nucleotide sequence of the Cclera sequence differs
from the published sequence for this transcript."
/product="CT25986"
/db_xref="FLYBASE:FBan0009054"
/db_xref="FLYBASE:FBgn0015075"
<69753..>72208
/gene="Ddx1"
/note="CG9054"
/map="79E1-79E2"
/db_xref="FLYBASE:FBan0009054"
/db_xref="FLYBASE:FBgn0015075"
join(69846..69861,69916..69967,70021..71690,71763..72208)
/gene="Ddx1"
/note="Ddx1 gene product; Nucleotide sequence of the
Cclera sequence differs from the published sequence for
this transcript"
/codon_start=1
/db_xref="FLYBASE:FBan0009054"
/db_xref="FLYBASE:FBgn0015075"
/protein_id="AAF51814.1"
/db_xref="GI:7296530"
```

alignment_scores:

Quality:	10.00	Length:	10
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-701-711-2 x AE003597 ..

Align seg 1/1 to: AE003597 from: 1 to: 295566

789	SerTyrAlaLysProLeuAsnLysLysGln	798
162573	TCTTATGCAAGCCACTTAACAAAAGCAA	162602

seq_name: gb_pat:AX078597

seq_documentation_block:
LOCUS AX078597 2019 bp DNA linear PAT 22-FEB-2001

DEFINITION Sequence 111 from Patent WO0107624.
ACCESSION AX078597
VERSION AX078597.1 GI:13158239
KEYWORDS
SOURCE Pseudomonas putida.
ORGANISM Pseudomonas putida
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 2019)
AUTHORS Fraser C.M., Venter C., Tuemmler B., Hoheisel J., Duenterhoeft A., Hilbert H., Timmis K.N., Moore E., Straetz M., Heim S., Nelson K.E., Hickey E. and Peterson J.
TITLE Dna sequences which are suited for specifically detecting Pseudomonas putida kt2440
JOURNAL Patent: WO 0107624-A 111 01-FEB-2001;
THE INSTITUTE FOR GENOMIC RESEARCH (US); QIAGEN GmbH (DE); Gesellschaft fuer Biotechnologische Forschung MBH (GBF) (DE); Deutsches Krebsforschungszentrum (DKFZ) (DE); Medizinische Hochschule Hannover (DE)
FEATURES
source Location/Qualifiers
1..2019
/organism="Pseudomonas putida"
/db_xref="taxon:303"
BASE COUNT 404 a 671 c 603 g 338 t 3 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AX078597 ..
Align seg 1/1 to: AX078597 from: 1 to: 2019

368 AspGluValLeuArgArgGluMetArg 376
|||||
250 GACGAAGTGTCTGCGTCGCGAAATGCGC 276
seq_name: gb_pr:AY062939

seq_documentation_block:
LOCUS AY062939 2418 bp DNA linear PRI 26-DEC-2001
DEFINITION Homo sapiens RETSDR2 (RETSDR2) gene, promoter region and partial cds.
ACCESSION AY062939
VERSION AY062939.1 GI:17978577
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2418)
AUTHORS Obeyesekere V.R., Saffery R.E. and Krozowski Z.S.
TITLE 17#HSDXI is a novel human microsomal dehydrogenase expressed in steroidogenic cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2418)
AUTHORS Obeyesekere V.R., Saffery R.E. and Krozowski Z.S.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Molecular Hypertension, Baker Institute, Commercial Rd, Melbourne, Vic 8008, Australia
FEATURES
source Location/Qualifiers
1..2418
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene <1..>2418
/gene="RETSDR2"
misc_feature <1..2415
/gene="RETSDR2"
/note="contains promoter region and 5' untranslated

region"
<2416..>2418
/gene="RETSDR2"
/product="RETSDR2"
2416..>2418
/gene="RETSDR2"
/codon_start=1
/product="RETSDR2"
/protein_id="AAL48315.1"
/db_xref="GI:17978578"
/translation="M"
BASE COUNT 731 a 514 c 494 g 678 t 1 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AY062939/rev ..
Align seg 1/1 to reverse of: AY062939 from: 1 to: 2418

46 IleGluSerLeuGlnSerValLeuPro 54
|||||
842 ATGAAAGTTTACAAAGTGTCTTCCA 816
seq_name: gb_pl:ZMZAG2

seq_documentation_block:
LOCUS ZMZAG2 7071 bp DNA linear PLN 29-SEP-1995
DEFINITION Z.mays zag2 gene.
ACCESSION X80206
VERSION X80206.1 GI:940880
KEYWORDS MADS-box protein; ZAG2 gene.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 7071)
AUTHORS Theissen G., Strater T., Fischer A. and Saedler H.
TITLE Structural characterization, chromosomal localization and phylogenetic evaluation of two pairs of AGAMOUS-like MADS-box genes from maize
JOURNAL Gene 156 (2), 155-166 (1995)
MEDLINE 95278740
REFERENCE 2 (bases 1 to 7071)
AUTHORS Theissen G.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-1994) G. Theissen, Max Planck Institut fuer Zuechtungsforschung, Carl von Linne Weg 10, D-50829 Koeln, Germany
REMARK Revised by [3]
REFERENCE 3 (bases 1 to 7071)
AUTHORS Theissen G.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1995) G. Theissen, Max Planck Institut fuer Zuechtungsforschung, Carl von Linne Weg 10, D-50829 Koeln, Germany
COMMENT On Aug 13, 1995 this sequence version replaced gi:854643.
Sequences overlapping with: R. Schmidt; Plant Cell, Vol. 5, 729-737, 1993.
FEATURES
source Location/Qualifiers
1..7071
/organism="Zea mays"
/sub_species="mays"
/db_xref="taxon:4577"
/chromosome="3"
/map="66"
/clone="EMBL4-II7b"
/clone_lib="EMBL4"
GC_signal 168..177

GC_signal /note="putative"
295..304
GC_signal /note="putative"
311..320
TATA_signal /note="putative"
392..399
TATA_signal /note="putative"
917..924
TATA_signal /note="putative"
1288..1297
TATA_signal /note="putative"
2058..2064
/note="putative"
Join(2133..2314,4276..4360,4755..4819,4959..5058,
5281..5322,5411..5452,5610..5900)
gene /gene="ZAG2"
2133..2314
exon /gene="ZAG2"
/number=1
Join(2133..2314,4276..4360,4755..4819,4959..5058,
5281..5322,5411..5452,5610..5900)
CDS /gene="ZAG2"
/codon_start=1
/protein_id="CAA56504.1"
/db_xref="GI:1001934"
/translation="MGRGRIETKRIENNTSRQVTFCKRRNGLLKAYELSVLCDAEVA
LIVFSRRLVEYANNVKATVERKKYKAHTVGSSSGPPLLHNNAOQFYQOFSAKLRNQ
IOMLQNTNRHLVGVDSVGNLSLKLQLESRLKLGSKTRAKSELLAAEISYMAKRET
BLONDHWTLRKIERGEQLOQVTVARVAAAATNLEAPFLFLEMDTKCFCTTGGPFA
TIDMKCFUPGSLQOMLEAQQRQMLATELNGLYQLAPPGSDAANNPHQF"
exon /gene="ZAG2"
/number=2
exon 4755..4819
/gene="ZAG2"
/number=3
exon 4959..5058
/gene="ZAG2"
/number=4
exon 5281..5322
/gene="ZAG2"
/number=5
exon 5411..5452
/gene="ZAG2"
/number=6
exon 5610..5900
/gene="ZAG2"
/number=7
3'UTR 5898..6197
polyA_signal 6060..6066
BASE COUNT 1813 a 1583 c 1526 g 2144 t 5 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x ZMZAG2/rev ..

Align seg 1/1 to reverse of: ZMZAG2 from: 1 to: 7071

489 ValSerGlnSerLeuSerGlyTyrTyr 497
|||||
2913 GTGAGTCAATCTTGTCTGGGTATAT 2887

seq_name: gb_pl:SPCC188

seq_documentation_block:
LOCUS SPCC188

35412 bp DNA linear 19-JAN-2000

DEFINITION S.pombe chromosome III cosmid c188.
ACCESSION AL049662
VERSION AL049662.1 GI:4678680
KEYWORDS 5s rRNA: alpha-amylase; GTP_cdc domain: LTR; nuclear protein;
phosphatase; pseudogene; ribonuclease; septin homolog; signal
recognition particle; srp54; tRNA; ubiquitin carboxyl-terminal
hydrolase.
SOURCE fission yeast.
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 35412)
AUTHORS Lyne,M., Rajandream,M.A., Barrell,B.G., Seeger,K. and Harris,D.
TITLE Direct Submission
JOURNAL Submitted (16-APR-1999) European Schizosaccharomycetes genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
During 1995 to 1996 about 66% of S. pombe chromosome I was
sequenced by the Sanger Centre. The sequencing of the S. pombe
genome is now being continued with funding from The European
Commission. Fourteen European sequencing laboratories, including
the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC25H2.01c. SP (S.
pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.

FEATURES
Source

1..35412
/organism="Schizosaccharomyces pombe"
/strain="972h"
/db_xref="taxon:4896"
/chromosome="III"
/map="IIIR"
/clone="cosmid c188"
complement(1..610)
/gene="SPCC188.01c"
/note="SPCC11E10.09c"
1..2559
/note="nominal overlap with cosmid SPCC11E10, EM:AL121783
S. pombe chromosome 3"
complement(1..610)
/partial
/gene="SPCC188.01c"
/note="SPCC188.01c, len>202,
SIMILARITY:Schizosaccharomycetes pombe, O13996, putative
alpha-amylase precursor, (491 aa), fasta scores: Opt: 833,
E():0, (60.9% identity in 184 aa)"
/codon_start=1
/label="SPCC188.01c
/product="putative alpha-amylase precursor"
/protein_id="CAB41221.1"
/db_xref="GI:4678681"
/translation="MNLICSLIPKKPSYKLWRKVIVQLTDRFALDEDFYAKAS
GNLYLGGTWKGTIRNLVDYIKSGCTAIWISPIVKNISSETDCGQAVHGVWADMTOLN
ENFGTEEDLKELVNAIHEKNMLCMVDIVVNMHGAGSKPVPFLQLQPFNSGKYHNWQ

gene
misc_feature

CDS

```

/rRNA
complement(1533..1696)
/notes="S. pombe 5s rRNA gene and flanks"
2318..2400
/notes="trna Lys anticodon CTT"
/product="trna-Lys"
join(3061..3084,3128..3223,3264..3477,3809..4563,
4616..4914,5004..5262)
/genes="SPCC188.02"
CDS
join(3061..3084,3128..3223,3264..3477,3809..4563,
4616..4914,5004..5262)
/genes="SPCC188.02"
/notes="SPCC188.02, len:547, SIMILARITY:Oryctolagus
cuniculus, Q28653, protein phosphatase pp2a0 b' subunit
gamma isoform., (586 aa), fasta scores: opt: 1779, E():0,
(52.3% identity in 507 aa)"
/codon_start=1
/label="SPCC188.02"
/product="putative protein phosphatase subunit"
/protein_id="CAB41222.1"
/db_xref="GI:4678682"
/translation="MKGIKMSLRGKSQDTOKSKKSKSKNSHDSKAPKSPST
DNGSVIGANDFLVPHKSGKKVPIDTTPTRDEITLLENVTRVKORSLSYHSENK
NLVRLPSTDPVKNWHSALAKLEQCCVFPDNPSTDLCKEYKREALQDLILIS
VKREALDESIPSIHMFVAVNFRPLPSPNPPEIMDEDEPALEVAWPHLHVID
FPLRFESPLTSVAKVYINQKFKLKLVLDFSDPRERDFLTKLHRIYKFLSLR
AFIRRSINLFTQYVYENQFGIAELLEILGSIINGFALPKKEHKIFLSVLPLH
KAKSLPYQIAYGVQVQVSEVWGLLRYKWPKNSSKEVLFNLNEDIEE
VMEPSEFLKIQVLPKHLATSISSONFOAERALLYFNNDYFVHLVEENVDLIPIY
PALFELSKSHNRVHSMVCNVILKFLMDINPLSFDEVDAYSESRKKEDEBIIEER
WTLLENIAKENAMLUKSNQPTTVHSTTERLKLKLSLDYING"
3085..3090
/misc_feature
/genes="SPCC188.02"
/notes="gtaaga, splice donor sequence"
3113..3127
/misc_feature
/genes="SPCC188.02"
/notes="gtaagg, splice donor sequence"
3249..3263
/misc_feature
/genes="SPCC188.02"
/notes="gtaagt, splice donor sequence"
3478..3483
/misc_feature
/genes="SPCC188.02"
/notes="gtaagt, splice donor sequence"
3792..3808
/misc_feature
/genes="SPCC188.02"
/notes="ctaacgtgtctattag, splice branch and acceptor"
4564..4569
/misc_feature
/genes="SPCC188.02"
/notes="gtaagt, splice donor sequence"
4598..4615
/misc_feature
/genes="SPCC188.02"
/notes="ctaacgtgacaatag, splice branch and acceptor"
4915..4920
/misc_feature
/genes="SPCC188.02"
/notes="gtaagt, splice donor sequence"
4989..5003
/misc_feature
/genes="SPCC188.02"
/notes="ctgacgtttgttag, splice branch and acceptor"
join(5588..7501,7542..7673,7734..7962,8011..8120,
8183..8425)
CDS
join(5588..7501,7542..7673,7734..7962,8011..8120,
8183..8425)
/genes="SPCC188.03"
/notes="SPCC188.03, len:874, SIMILARITY:Saccharomyces
cerevisiae, Q06680, chromosome iv cosmid 9798., (1051 aa),
fasta scores: opt: 406, E():2.4e-18, (30.0% identity in
949 aa)"
/codon_start=1
/label="SPCC188.03"

```

```

/product="hypothetical protein"
/protein_id="CAB41223.1"
/db_xref="GI:4678683"
/translation="MSCIQLISSOTSIAHRKLCNKLETLTFOEGFTDILRALNII
LVKGNNSADRVLRFLVTFVNYLQKQDEIDIVQILKHLGLDAKQDLYVRCCQ
LIARVGNCKEIDDDLYNTLTKLLSRDLDSIVRLAEVAVLSLQEDTGDSENDVR
NIFLLQNDSPSEVRVLLNIEVNSSTLPFLERARDVDAANRKCVCYARVLPIKIGD
FYLISIKRRVRIILKWLNDRESEVKAADMEYLAQWIENADNLLLELLERLDVSNNSD
VAVLAIKFFDVRVDSLSOLEPPEQFWLELTAESSLARTFNEICTEKNTDLDKMP
EVOLVYIYERQVRSRDKSSYDESCFIIETOLLYIGLSODMVDIEGRKLLKSLTNSL
SMAAPDLSLSLHIELLRKLCSEDFCSLLVEIITEVFEQHSQNTQEQGSNAPE
LNKNDYEGEITVSQSPSPSPNPNEPEPDMDGCKEAFNELKCLSTVQCCLFENI
TSLNLYNVMVMDLKTLLIPAVNSHDLPIREKGLCCLSLVCLLNADLAFENVPLYLHC
YKGSVYKLCVTAIRTDMLIQHKAKFTEYDAISILFEALGEFENAELOLIGAEA
IARGLVILHYRDELFLKPLTIQVFEPTVDNHALRQVLGVFFVYAFGAHENGWRAT
IFCDALLSLEIYRDLDEVDOLSIGIAQMDLDWTDNKLRYKRTQTDGDIYALNHN
VHLHLANMIFESLPNASEGKERKFMISLLGKLKIPDLPSSDYQRTKRLKETYESHGF
TWDSISLSILAKFERMLQNEARSKFEETEERLMEAEENEHAGAEASIGELIPT
VEANMEDEBEVYKQEDL"
7502..7507
/misc_feature
/genes="SPCC188.03"
/notes="gtatgt, splice donor sequence"
7528..7541
/misc_feature
/genes="SPCC188.03"
/notes="ctacaacaactag, splice branch and acceptor"
7674..7679
/misc_feature
/genes="SPCC188.03"
/notes="gtaagg, splice donor sequence"
7719..7733
/misc_feature
/genes="SPCC188.03"
/notes="ctaacagcatttag, splice branch and acceptor"
7963..7968
/misc_feature
/genes="SPCC188.03"
/notes="gtgggg, splice donor sequence"
7995..8010
/misc_feature
/genes="SPCC188.03"
/notes="ctacaacaactag, splice branch and acceptor"
8121..8126
/misc_feature
/genes="SPCC188.03"
/notes="gtaagt, splice donor sequence"
8169..8182
/misc_feature
/notes="ctaacatgtatag, splice branch and acceptor"
complement(join(8422..8564,8629..9042,9091..9250))
/genes="SPCC188.04c"
CDS
complement(join(8422..8564,8629..9042,9091..9250))
/genes="SPCC188.04c"
/notes="SPCC188.04c, len:237, SIMILARITY:Saccharomyces
cerevisiae, Q12234, chromosome xv reading frame orf
yor216c., (484 aa), fasta scores: opt: 163, E():0.023,
(26.8% identity in 194 aa)"
/codon_start=1
/label="SPCC188.04c"

```

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x SPCC188 ..

Align seg 1/1 to: SPCC188 from: 1 to: 35412

237 LeuasnAlaGlyPheValArgPheGlu 245

|||||

27972 TTAACGCGGGCTTTGTAAGATTGAA 27998

seq_name: gb_pl:SPAC8A4

seq_documentation_block:

LOCUS SPAC8A4

DEFINITION S.pombe chromosome I cosmid c8A4.

38141 bp DNA linear

PLN 04-NOV-1995

homology to SPAC8A4.02c but with more (about 14) 36 aa repeats"

gene complement(9025..9558)
/gene="SPAC8A4.04c"

CDS complement(9025..9558)
/gene="SPAC8A4.04c"

/note="Protein sequence is in conflict with the conceptual translation; SPAC8A4.04c, unknown orf, len: 177, possibly spliced but the three separate reading frames in this region are only separated by single stop codons and all share high similarity to SPAC8A4.02c. Therefore these 3 orfs could be a pseudogene"

/codon_start=1
/product="unknown"

/protein_id="CAA91514.1"

/db_xref="GI:1326031"

/translation="FAFDSVSRVVEVIEPIAGTLTSTVYSGSGFTTTTGTASGSVSGTVEVQPTAGTITTVYSGTGGTTLTGTASGSVGTVEVQPTAGTITTVYTFEGSSTFTVLNLSAGSVSGTVDVVKAGPNTTYSGTPTTVYSSASSAATVVVPIPTAVCSGI RGLQYAVNYNTIASSKS"

gene complement(9559..11121)
/gene="SPAC8A4.05c"

CDS complement(9559..11121)
/gene="SPAC8A4.05c"

/note="SPAC8A4.05c, unknown orf, len: 520, possibly spliced to downstream orf but the three separate reading frames in this region are only separated by single stop codons and all share high similarity to SPAC8A4.02c. Therefore these 3 orfs could be a pseudogene"

/codon_start=1
/product="unknown"

/protein_id="CAA91515.1"

/db_xref="GI:1052536"

/db_xref="SWISS-PROT:Q09881"

/translation="MFSEFRSTLVFLFISFTVVLSEPRVSIGLAKRTISQTSSSSLYTCPDYITIFSEGSEYTTIYPSKSSSVANHSITRTIDSGTIATYFTFLPDGEI VIRDIEPVAKTLTTITSGSLELTTLTATASGTVEVIEVPLAGTATVTVYSGSVE HNTLTATASGTVEVIEVPLAGTATVTVYSGSVEVIEVPLAGTATVTVYSGSVE AGTVTTIYSGSVENLTATASETLPGTIEVVEPLAGTATVTVYSGSQAQETTLATA SCTVSGTVEVIEVPLAGTATVTVYSGSEATTVASGTVEVETAGTVYQTSMDLPL TSLTVTATASGTVEVIEVPLAGTATVTVYSGSQYTTTATAGCVSGVGVVEVIEP ASGAVGTVEVVDPAAGTATVTVYSGSQTTLATATATGIPGTVEIVDPAAGTATV TYSGMEYRSTL"

gene join(13149..13177,13221..14772,14821..14913)
/gene="SPAC8A4.06"

CDS join(13149..13177,13221..14772,14821..14913)
/note="SPAC8A4.06, similar to BX42_DROME P39736 puff specific protein BX42, (46.0% identity in 409 aa overlap)"

/codon_start=1
/product="unknown"

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x SPAC8A4

Align seg 1/1 to: SPAC8A4 from: 1 to: 38141

237 LeuAsnAlaGlyPheValArgPheGlu 245

11145 TTAACGGCGGCTTGTGAAGATTGAA 11171

seq_name: gb_pl:SPCC584

seq_documentation_block:

LOCUS SPCC584

DEFINITION S.pombe chromosome III cosmid c584.

ACCESSION AL032824

VERSION KEYWORDS

AL032824.2 GI:4456829

amino-acid permease; Chromatin binding; FAD/NAD-binding domain; LTR; metal-binding regulatory protein; replication factor-a protein 2; ribonuclease; sec1 family; secretory protein; septin homolog; snw family nuclear protein; sulfite reductase; syntaxin binding protein.

SOURCE

ORGANISM

Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 41803)
Direct Submission
Submitted (29-OCT-1998) European Schizosaccharomyces genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrellesanger.ac.uk
On Mar 21, 1999 this sequence version replaced gi:3819700.

Notes:

Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/S_pombe/)

During 1995 to 1996 about 66% of S. pombe chromosome 1 was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project. Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in Pombase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites. CDS are numbered using the following system eg SPCC35H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. c584 overlaps c188 at its 5' end, and c1753 at its 3' end.

FEATURES

source

1..41803

/organism="Schizosaccharomyces pombe"

/strain="972h"

/db_xref="taxon:4896"

/chromosome="III"

/map="IIIR"

/clone="cosmid c584"

/complement(1..649)

/gene="SPCC584.07c"

/complement(1..649)

/partial

/gene="SPCC584.07c"

/note="SPCC584.07c, len:<216: contains 36 aa repeat"

/codon_start=1

/label="SPCC584.07c"

/product="hypothetical repeat containing protein"

/protein_id="CAB37420.1"

/db_xref="GI:4456830"

/translation="MFSEFRSTLVFLFISFTVVLSEPRVSIGLAKRTISQTSSSSLY TCPDYITIFSEGSEYTTIYPSKSSSVANHSITRTIDSGTIATYFTFLPDGEI VIRDIEPVAKTLTTITSGSLELTTLTATASGTVEVIEVPLAGTATVTVYSGSVE HNTLTATASGTVEVIEVPLAGTATVTVYSGSQAQETTLATAAGCVSGVGVVEVIEP ASGAVGTVEVVDPAAGTATVTVYSGSQTTLATATATGIPGTVEIVDPAAGTATV TYSGMEYRSTL"

1..8113

/note="overlap with c188 S. pombe chromosome 3"

2047..2159

/note="small conserved intergenic region, possible rRNA, 2 3532 23644 c188 S. pombe chromosome 3"

join(2677..2705,2749..4300,4349..4441)

misc_feature

misc_feature

gene

KEYWORDS

HTG; HTGS_PHASE0.

SOURCE

house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 64341)

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Mus musculus, clone RP24-319020

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 64341)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Bouckgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Collangelo,M., Collins,S., Collamore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Glade,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
McLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Taldamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

TITLE

Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome

JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20053

Center clone name: 319_Q_20

* NOTE: This record contains 82 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

* 1 676: contig of 676 bp in length

* 677 776: contig of 100 bp

* 777 1447: contig of 671 bp in length

* 1448 1547: gap of 100 bp

* 1548 2204: contig of 657 bp in length

* 2205 2304: gap of 100 bp

* 2305 2971: contig of 667 bp in length

* 2972 3071: gap of 100 bp

* 3072 3689: contig of 618 bp in length

* 3690 3789: gap of 100 bp

* 3790 4489: contig of 700 bp in length

* 4490 4589: gap of 100 bp

* 4590 5287: contig of 698 bp in length

* 5288 5387: gap of 100 bp

* 5388 6073: contig of 686 bp in length

6074 6173: gap of 100 bp
6174 6843: contig of 670 bp in length
6844 6943: gap of 100 bp
6944 7584: contig of 641 bp in length
7585 7684: gap of 100 bp
7685 8386: contig of 702 bp in length
8387 8486: gap of 100 bp
8487 9164: contig of 678 bp in length
9165 9264: gap of 100 bp
9265 9947: contig of 683 bp in length
9948 10047: gap of 100 bp
10048 10755: contig of 708 bp in length
10756 10855: gap of 100 bp
10856 11554: contig of 699 bp in length
11555 11654: gap of 100 bp
11655 12351: contig of 697 bp in length
12352 12451: gap of 100 bp
12452 13145: contig of 694 bp in length
13146 13245: gap of 100 bp
13246 13906: contig of 661 bp in length
13907 14006: gap of 100 bp
14007 14690: contig of 684 bp in length
14691 14790: gap of 100 bp
14791 15484: contig of 694 bp in length
15485 15584: gap of 100 bp
15585 16279: contig of 695 bp in length
16280 16379: gap of 100 bp
16380 17068: contig of 689 bp in length
17069 17168: gap of 100 bp
17169 17855: contig of 687 bp in length
17856 17955: gap of 100 bp
17956 18636: contig of 681 bp in length
18637 18736: gap of 100 bp
18737 19424: contig of 688 bp in length
19425 19524: gap of 100 bp
19525 20243: contig of 719 bp in length
20244 20343: gap of 100 bp
20344 21059: contig of 716 bp in length
21060 21159: gap of 100 bp
21160 21865: contig of 706 bp in length
21866 21965: gap of 100 bp
21966 22664: contig of 699 bp in length
22665 22764: gap of 100 bp
22765 23440: contig of 676 bp in length
23441 23540: gap of 100 bp
23541 24221: contig of 681 bp in length
24222 24321: gap of 100 bp
24322 25010: contig of 689 bp in length
25011 25110: gap of 100 bp
25111 25812: contig of 702 bp in length
25813 25912: gap of 100 bp
25913 26615: contig of 703 bp in length
26616 26715: gap of 100 bp
26716 27390: contig of 675 bp in length
27391 27490: gap of 100 bp
27491 28202: contig of 712 bp in length
28203 28302: gap of 100 bp
28303 28994: contig of 692 bp in length
28995 29094: gap of 100 bp
29095 29772: contig of 678 bp in length
29773 29872: gap of 100 bp
29873 30580: contig of 708 bp in length
30581 30680: gap of 100 bp
30681 31369: contig of 689 bp in length
31370 31469: gap of 100 bp
31470 32134: contig of 665 bp in length
32135 32234: gap of 100 bp
32235 32923: contig of 689 bp in length
32924 33023: gap of 100 bp
33024 33712: contig of 689 bp in length
33713 33812: gap of 100 bp
33813 34511: contig of 699 bp in length
34512 34611: gap of 100 bp

```

* 34612 35283: contig of 672 bp in length
* 35284 35383: gap of 100 bp
* 35384 36105: contig of 722 bp in length
* 36106 36205: gap of 100 bp
* 36206 36884: contig of 679 bp in length
* 36885 36984: gap of 100 bp
* 36985 37658: contig of 674 bp in length
* 37659 37758: gap of 100 bp
* 37759 38414: contig of 656 bp in length
* 38415 38514: gap of 100 bp
* 38515 39196: contig of 682 bp in length
* 39197 39296: gap of 100 bp
* 39297 40003: contig of 707 bp in length
* 40004 40103: gap of 100 bp
* 40104 40772: contig of 669 bp in length
* 40773 40872: gap of 100 bp
* 40873 41560: contig of 688 bp in length
* 41561 41660: gap of 100 bp
* 41661 42309: contig of 649 bp in length
* 42310 42409: gap of 100 bp
* 42410 43115: contig of 706 bp in length
* 43116 43215: gap of 100 bp
* 43216 43908: contig of 693 bp in length
* 43909 44008: gap of 100 bp
* 44009 44702: contig of 694 bp in length
* 44703 44802: gap of 100 bp
* 44803 45498: contig of 696 bp in length
* 45499 45598: gap of 100 bp
* 45599 46281: contig of 683 bp in length
* 46282 46381: gap of 100 bp
* 46382 47067: contig of 686 bp in length
* 47068 47167: gap of 100 bp
* 47168 47830: contig of 663 bp in length
* 47831 47930: gap of 100 bp
* 47931 48611: contig of 681 bp in length
* 48612 48711: gap of 100 bp
* 48712 49396: contig of 685 bp in length
* 49397 49496: gap of 100 bp
* 49497 50186: contig of 690 bp in length
* 50187 50286: gap of 100 bp
* 50287 50956: contig of 670 bp in length
* 50957 51056: gap of 100 bp
* 51057 51767: contig of 711 bp in length
* 51768 51867: gap of 100 bp
* 51868 52565: contig of 698 bp in length
* 52566 52665: gap of 100 bp
* 52666 53364: contig of 699 bp in length
* 53365 53464: gap of 100 bp
* 53465 54141: contig of 677 bp in length
* 54142 54241: gap of 100 bp

```

alignment_scores:

```

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-09-701-711-2 x AC102899/rev ..

```

```

Align seg 1/1 to reverse of: AC102899 from: 1 to: 64341

```

```

607 LeuThrValGlyPheGlyAspLysThr 615
|||||

```

```

48914 CTCACGGTTGGTTGGGACAGACG 48888

```

```

seq_name: gb_htg:AC095320

```

```

seq_documentation_block:

```

```

LOCUS AC095320 94948 bp DNA linear HTG 20-DEC-2001
DEFINITION Rattus norvegicus clone CH230-141D12, *** SEQUENCING IN PROGRESS
*** 57 unordered pieces.
ACCESSION AC095320
VERSION AC095320.2 GI:17956648

```

KEYWORDS

```

SOURCE HTG; HTGS_PHASE1.

```

ORGANISM

```

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

REFERENCE

AUTHORS

```

1 (bases 1 to 94948)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
Benton,J., Binage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsfi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkwo,S.,
Ogih,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshkari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,F., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

```

Direct Submission

```

Unpublished

```

```

2 (bases 1 to 94948)

```

```

Worley,K.C.

```

JOURNAL

REFERENCE

```

Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15625874.

```

COMMENT

```

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFXY
Center clone name: CH230-141D12
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 67588 bases at least Q40
Consensus quality: 73195 bases at least Q30
Consensus quality: 77056 bases at least Q20
Estimated insert size: 31978; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-ep estimation
Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation
-----

```

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 57 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3273: contig of 3273 bp in length
* 3274: gap of unknown length
* 3374: contig of 2037 bp in length
* 5411: gap of unknown length
* 5511: contig of 2134 bp in length
* 7645: gap of unknown length
* 7744: contig of 1012 bp in length
* 8756: gap of unknown length
* 8857: contig of 2117 bp in length
* 10973: gap of unknown length
* 10974: contig of 2257 bp in length
* 13330: gap of unknown length
* 13331: contig of 1485 bp in length
* 13431: gap of unknown length
* 14915: contig of 2683 bp in length
* 15015: gap of unknown length
* 17698: contig of 1741 bp in length
* 17799: gap of unknown length
* 19539: contig of 2878 bp in length
* 19540: gap of unknown length
* 19640: contig of 1821 bp in length
* 22518: gap of unknown length
* 22617: contig of 1933 bp in length
* 24438: gap of unknown length
* 24539: contig of 1132 bp in length
* 26471: gap of unknown length
* 26571: contig of 1132 bp in length
* 26572: gap of unknown length
* 27704: contig of 2090 bp in length
* 27803: gap of unknown length
* 29893: contig of 2268 bp in length
* 29994: gap of unknown length
* 32361: contig of 1456 bp in length
* 32362: gap of unknown length
* 32363: contig of 2306 bp in length
* 33917: gap of unknown length
* 36223: contig of 1422 bp in length
* 36323: gap of unknown length
* 36324: contig of 1372 bp in length
* 37745: gap of unknown length
* 37846: contig of 1372 bp in length
* 39217: gap of unknown length
* 39317: contig of 1464 bp in length
* 40781: gap of unknown length
* 40881: contig of 1985 bp in length
* 42866: gap of unknown length
* 42967: contig of 1441 bp in length
* 44408: gap of unknown length
* 44507: contig of 1090 bp in length
* 45597: gap of unknown length
* 45697: contig of 1235 bp in length
* 45698: gap of unknown length
* 46933: contig of 1008 bp in length
* 47033: gap of unknown length
* 48040: contig of 1568 bp in length
* 48140: gap of unknown length
* 49708: contig of 1294 bp in length
* 49709: gap of unknown length
* 49809: contig of 1629 bp in length
* 51437: gap of unknown length
* 51537: contig of 1804 bp in length
* 53341: gap of unknown length
* 53441: contig of 1294 bp in length
* 54735: gap of unknown length
* 54835: contig of 1435 bp in length
* 56270: gap of unknown length
* 56370: contig of 1048 bp in length
* 57418: gap of unknown length
* 57518: contig of 1102 bp in length
* 58620: gap of unknown length

* 58621: gap of unknown length
* 58721: contig of 1356 bp in length
* 60077: gap of unknown length
* 60176: contig of 2197 bp in length
* 62374: gap of unknown length
* 62474: contig of 1452 bp in length
* 63926: gap of unknown length
* 64026: contig of 1601 bp in length
* 65627: gap of unknown length
* 65727: contig of 1917 bp in length
* 67644: gap of unknown length
* 67743: contig of 1406 bp in length
* 69149: gap of unknown length
* 69249: contig of 1094 bp in length
* 69250: gap of unknown length
* 70344: contig of 1876 bp in length
* 70444: gap of unknown length
* 72120: contig of 1199 bp in length
* 72220: gap of unknown length
* 73419: contig of 1236 bp in length
* 73519: gap of unknown length
* 74755: contig of 1259 bp in length
* 74855: gap of unknown length
* 76114: contig of 1847 bp in length
* 76214: gap of unknown length
* 78061: contig of 1413 bp in length
* 78161: gap of unknown length
* 79574: contig of 1117 bp in length
* 79674: gap of unknown length
* 80791: contig of 1288 bp in length
* 80891: gap of unknown length
* 82179: contig of 1009 bp in length
* 82279: gap of unknown length
* 83288: contig of 1284 bp in length
* 83388: gap of unknown length
* 84672: contig of 1431 bp in length
* 84772: gap of unknown length
* 86202: contig of 1183 bp in length
* 86303: gap of unknown length
* 87486: contig of 1123 bp in length
* 87586: gap of unknown length
* 88709: contig of 1351 bp in length
* 88809: gap of unknown length
* 90160: contig of 1225 bp in length
* 90259: gap of unknown length
* 90260: contig of 1106 bp in length
* 91484: gap of unknown length
* 91585: contig of 1106 bp in length
* 91585: gap of unknown length
* 92691: contig of 1025 bp in length
* 92791: gap of unknown length
* 93815: contig of 1025 bp in length
* 93816: gap of unknown length

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AC095320/rev ..

Align seg 1/1 to reverse of: AC095320 from: 1 to: 94948

679 AlaArgGlyGlnGlnThrLeu 687

|||||

29093 GCCAAGAGGGCAGCAGACATTG 29067

seq_name: gb_pr:AL359454

seq_documentation_block:

LOCUS AL359454

DEFINITION Human DNA sequence from clone RP11-97E23 on chromosome 13, complete

sequence.

ACCESSION AL359454

VERSION AL359454.7 GI:14272304

KEYWORDS HTG.

human.. Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 101830)
Smith, M.
Direct Submission
Submitted (30-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
On May 31, 2001 this sequence version replaced gi:13398783.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: EM: EMBL; SW:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr13>
RP11-97E23 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-97E23 It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-97E23 is at 101830 in this
sequence. The true left end of clone RP11-161P17 is at 78762 in
this sequence. The true right end of clone RP11-57H24 is at 100 in
this sequence.

FEATURES
source
1..101830
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-97E23"
/clone_lib="RPC1-11.1"
671..980
repeat_region
/note="AluX repeat: matches 1..310 of consensus"
2541..2808
repeat_region
/note="AluJo repeat: matches 34..296 of consensus"
3048..3084
repeat_region
/note="MIR repeat: matches 156..192 of consensus"
3577..3713
repeat_region
/note="MIR repeat: matches 4..143 of consensus"
4567..4753
repeat_region
/note="HV3 repeat: matches 1..99 of consensus"
4756..5060
repeat_region
/note="AluSq repeat: matches 1..305 of consensus"
5291..5428
repeat_region
/note="LTR16C repeat: matches 248..381 of consensus"
5429..5740
repeat_region
/note="AluX repeat: matches 1..312 of consensus"
5741..5896
repeat_region
/note="LTR16C repeat: matches 79..248 of consensus"
6321..7101
repeat_region
/note="MIR repeat: matches 23..215 of consensus"
7357..7669
repeat_region
/note="AluSq repeat: matches 1..312 of consensus"

```

/note="L2 repeat: matches 2069. .2419 of consensus"
32005. .32317
/note="AlusX repeat: matches 1. .307 of consensus"
32812. .32940
/note="FLAM_A repeat: matches 16. .140 of consensus"
33242. .33404
/note="MIR repeat: matches 84. .261 of consensus"
33747. .34249
/note="CpG island"
/evidence=not_experimental
35100. .35458
/note="MER47A repeat: matches 3. .366 of consensus"
35487. .35554
/note="MIR repeat: matches 67. .137 of consensus"
35754. .35850
/note="L2 repeat: matches 2005. .2109 of consensus"
36054. .36356
/note="AluJb repeat: matches 1. .300 of consensus"
36447. .36511
/note="L2 repeat: matches 2679. .2743 of consensus"
37117. .37963
/note="MIR repeat: matches 2. .262 of consensus"
38176. .38482
/note="AlusP repeat: matches 1. .308 of consensus"
39199. .39308
/note="L1MD2 repeat: matches 6014. .6122 of consensus"
41473. .41584
/note="MIR repeat: matches 23. .155 of consensus"
41678. .41915
/note="MER20 repeat: matches 1. .214 of consensus"
42119. .42278
/note="FAM repeat: matches 1. .168 of consensus"
43498. .43669
/note="L2 repeat: matches 742. .919 of consensus"
43708. .43769
/note="L2 repeat: matches 1724. .1783 of consensus"
43807. .44242
/note="L2 repeat: matches 958. .1327 of consensus"
44243. .44504
/note="AluJo repeat: matches 49. .306 of consensus"
44505. .44698
/note="L2 repeat: matches 1327. .1490 of consensus"
44699. .44993
/note="AluJb repeat: matches 3. .293 of consensus"
44994. .45418
/note="L2 repeat: matches 1490. .2173 of consensus"
45638. .45929
/note="Alusq repeat: matches 1. .292 of consensus"

alignment_scores:
  Quality: 9.00 Length: 9
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AL359454/rev ..
Align seg 1/1 to reverse of: AL359454 from: 1 to: 101830

502 TyrAspAsnLysAsnIleSerAsnTyr 510
|||||
57836 TATGATAACAAAAATATCTCAAAATTAT 57810

seq_name: gb_htg:AC094530
seq_documentation_block:
LOCUS AC094530
DEFINITION Rattus norvegicus clone CH230-4L15, *** SEQUENCING IN PROGRESS ***,
53 unordered pieces.
ACCESSION AC094530
VERSION AC094530.3 GI:17941265
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.

```

```

ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
1 (bases 1 to 106997)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbacia,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,
Hollins,B., Homsif,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okunou,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Perry,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu.L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoostari,N.,
Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Telford,B., Thomas,N.,
Tang,H., Tansey,J., Taylor,C., Taylor,V., Villalon,D., Vinson,R.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Williamson,C.,
Watlington,S., Williams,G., Williams,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 106997)
Worley,K.C.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15799323.
-----
Center: Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
-----
Project Information
Center project name: GAVE
Center clone name: CH230-4L15
-----
Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 78042 bases at least Q40
Consensus quality: 86842 bases at least Q30
Consensus quality: 93110 bases at least Q20
Estimated insert size: 79937; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 0.9x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 4485: contig of 4485 bp in length
* 4486 4585: gap of unknown length
* 4586 10783: contig of 6198 bp in length
* 10784 10883: gap of unknown length
* 10884 13167: contig of 2284 bp in length
* 13168 13267: gap of unknown length
* 13268 16560: contig of 3293 bp in length
* 16561 16660: gap of unknown length
* 16661 19562: contig of 2902 bp in length
* 19563 19662: gap of unknown length
* 19663 21931: contig of 2289 bp in length
* 21932 22031: gap of unknown length
* 22032 25291: contig of 3260 bp in length
* 25292 25391: gap of unknown length
* 25392 26888: contig of 1497 bp in length
* 26889 29002: contig of 2014 bp in length
* 29003 31586: contig of 2484 bp in length
* 31587 31686: gap of unknown length
* 31687 34451: contig of 2765 bp in length
* 34452 34551: gap of unknown length
* 34552 37275: contig of 2724 bp in length
* 37276 37375: gap of unknown length
* 37376 39363: contig of 1988 bp in length
* 39364 41942: contig of 2479 bp in length
* 41943 42042: gap of unknown length
* 42043 44545: contig of 2503 bp in length
* 44546 46144: gap of unknown length
* 46145 46314: gap of unknown length
* 46315 48423: contig of 1909 bp in length
* 48424 48523: gap of unknown length
* 48524 50116: contig of 1593 bp in length
* 50117 50216: gap of unknown length
* 50217 52349: contig of 2033 bp in length
* 52350 52349: gap of unknown length
* 52350 54385: contig of 2036 bp in length
* 54386 54485: gap of unknown length
* 54486 56136: contig of 1651 bp in length
* 56137 56236: gap of unknown length
* 56237 57268: contig of 1032 bp in length
* 57269 57368: gap of unknown length
* 57369 59226: contig of 1858 bp in length
* 59227 59326: gap of unknown length
* 59327 60774: contig of 1448 bp in length
* 60775 60874: gap of unknown length
* 60875 63429: contig of 2555 bp in length
* 63430 63529: gap of unknown length
* 63530 65824: contig of 2295 bp in length
* 65825 65924: gap of unknown length
* 65925 67471: contig of 1547 bp in length
* 67472 67571: gap of unknown length
* 67572 69443: contig of 1872 bp in length
* 69444 69543: gap of unknown length
* 69544 71798: contig of 2255 bp in length
* 71799 71898: gap of unknown length
* 71899 73517: contig of 1619 bp in length
* 73518 73617: gap of unknown length
* 73618 75617: contig of 2000 bp in length
* 75618 77253: contig of 1536 bp in length
* 77254 77353: gap of unknown length
* 77354 79220: contig of 1867 bp in length

* 79221 79320: gap of unknown length
* 79321 80763: contig of 1443 bp in length
* 80764 82206: contig of 1343 bp in length
* 82207 82306: gap of unknown length
* 82307 84307: contig of 2001 bp in length
* 84308 84407: gap of unknown length
* 84408 85870: contig of 1463 bp in length
* 85871 85970: gap of unknown length
* 85971 87046: contig of 1076 bp in length
* 87047 87146: gap of unknown length
* 87147 88524: contig of 1378 bp in length
* 88525 88624: gap of unknown length
* 88625 89791: contig of 1167 bp in length
* 89792 89891: gap of unknown length
* 89892 91038: contig of 1147 bp in length
* 91039 91138: gap of unknown length
* 91139 92175: contig of 1037 bp in length
* 92176 92275: gap of unknown length
* 92276 93492: contig of 1217 bp in length
* 93493 93592: gap of unknown length
* 93593 94771: contig of 1179 bp in length
* 94772 94871: gap of unknown length
* 94872 95994: contig of 1123 bp in length
* 95995 96094: gap of unknown length
* 96095 97483: contig of 1389 bp in length
* 97484 97583: gap of unknown length
* 97584 98809: contig of 1226 bp in length
* 98810 98909: gap of unknown length
* 98910 100242: contig of 1333 bp in length
* 100243 100342: gap of unknown length
* 100343 101569: contig of 1227 bp in length
* 101570 101669: gap of unknown length
* 101670 103028: contig of 1359 bp in length
* 103029 103128: gap of unknown length
* 103129 104159: contig of 1031 bp in length
* 104160 104259: gap of unknown length
* 104260 105435: contig of 1176 bp in length
* 105436 105535: gap of unknown length
* 105536 106997: contig of 1462 bp in length.

FEATURES

source
1..106997
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-4L15"

BASE COUNT 30338 a 20613 c 20041 g 30563 t 5442 others
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AC094530 ..

Align seg 1/1 to: AC094530 from: 1 to: 106997

632 LysSerValLeuArgGlyTyrAlaLys 640

|||||

101456 AAATCTGTACTAAGAGATATGCAAAA 101482

seq_name: gb_htg:AC087596

seq_documentation_block:

LOCUS AC087596 109210 bp DNA linear HTG 12-JAN-2001
DEFINITION Oryza sativa chromosome 1 clone OSJNBa0055C02, *** SEQUENCING IN
PROGRESS ***, 3 ordered pieces.
ACCESSION AC087596
VERSION AC087596.1 GI:12084798
KEYWORDS HTG; HTGS-PHASE2.
SOURCE Oryza sativa.
ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 109210)
 AUTHORS Lee, J.-S., Hahn, J.-H., Lee, M.-C., Yoon, U.-H., Yun, D.-W., Kim, H.-I. and Eun, M.-Y.

TITLE Oryza sativa BAC OSJNBa0055C02 genomic sequence

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 109210)

AUTHORS Hahn, J.-H. and Eun, M.-Y.

TITLE Direct Submission

JOURNAL Submitted (12-JAN-2001) Rice Genome Sequencing Project, National

Institute of Agricultural Science and Technology (NIAT), RDA, 249

Seodun-dong, Suwon 441-707, Korea (E-mail: jhahn@rda.go.kr,

Tel: 82-31-290-0309, Fax: 82-31-290-0308)

* NOTE: This is a 'working draft' sequence. It currently

* consists of 3 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 25874: contig of 25874 bp in length

* 25875 25974: gap of unknown length

* 25975 105215: contig of 79241 bp in length

* 105216 105315: gap of unknown length

* 105316 109210: contig of 3895 bp in length.

FEATURES Location/Qualifiers

source

1..109210

/organism="Oryza sativa"

/cultivar="Nipponbare"

/db_xref="taxon:4530"

/chromosome="1"

/clone="OSJNBa0055C02"

BASE COUNT 31001 a 23200 c 23520 g 31287 t 202 others

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AC087596 ..

Align seg 1/1 to: AC087596 from: 1 to: 109210

514 SerTyrGlyGlySerLeuSerTyrGly 522

|||||

15087 TCCTACGGGGATCATTTGCTACGGT 15113

seq_name: gb_htg:LMFLCHR16_09

seq_documentation_block:

WPCOMMENT

Sequence split into 11 fragments LOCUS LMFLCHR16 Accession AL499619

Fragment Name Begin End

LMFLCHR16_00 1 110000

LMFLCHR16_01 100001 210000

LMFLCHR16_02 200001 310000

LMFLCHR16_03 300001 410000

LMFLCHR16_04 400001 510000

LMFLCHR16_05 500001 610000

LMFLCHR16_06 600001 710000

LMFLCHR16_07 700001 810000

LMFLCHR16_08 800001 910000

LMFLCHR16_09 900001 1010000

LMFLCHR16_10 1000001 1030105

Continuation (10 of 11) of LMFLCHR16 from base 900001 (AL499619 Leishmania major chromos

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x LMFLCHR16_09/rev ..

Align seg 1/1 to reverse of: LMFLCHR16_09 from: 1 to: 110000

287 GlnAlaGluLeuGluAlaLeuLeuLys 295

|||||

76655 CAGGCCGAGCTTGAGCGCTGCTGAAG 76629

seq_name: gb_htg:AC016033

seq_documentation_block:

LOCUS AC016033

DEFINITION Homo sapiens clone RP11-13024, *** SEQUENCING IN PROGRESS ***, 11

unordered pieces.

ACCESSION AC016033

VERSION AC016033.5 GI:17977644

KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 117693)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome, clone RP11-13024

REFERENCE 2 (bases 1 to 117693)

AUTHORS

JOURNAL

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhgalter, B.,

Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,

Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,

Lehoczky, J., Liu, C., Locke, K., Macdonald, P., Marquis, N.,

McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,

Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,

Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,

Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL

COMMENT

On Dec 23, 2001 this sequence version replaced gi:15290989.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3376

Center clone name: 13_O_24

* NOTE: This is a 'working draft' sequence. It currently

* consists of 11 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 19406: contig of 19406 bp in length

* 19407 19506: gap of 100 bp

```

* 19507 35142: contig of 15636 bp in length
* 35143 35242: gap of 100 bp
* 35243 40193: contig of 4951 bp in length
* 40194 40293: gap of 100 bp
* 40294 52911: contig of 12618 bp in length
* 52912 53011: gap of 100 bp
* 53012 57184: contig of 4173 bp in length
* 57185 57284: gap of 100 bp
* 57285 63653: contig of 6369 bp in length
* 63654 63753: gap of 100 bp
* 63754 105066: contig of 41313 bp in length
* 105067 105166: gap of 100 bp
* 105167 107339: contig of 2173 bp in length
* 107340 107439: gap of 100 bp
* 107440 112075: contig of 4636 bp in length
* 112076 112175: gap of 100 bp
* 112176 115003: contig of 2828 bp in length
* 115004 115103: gap of 100 bp
* 115104 117693: contig of 2590 bp in length.
FEATURES
      source
      1. .117693
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="Rp11-13024"
      /clone_lib="RPC1-11 Human Male BAC"
BASE COUNT  32428 a 26676 c 26081 g 31446 t 1062 others
ORIGIN

```

```

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
US-09-701-711-2 x AC016033/rev ..
Align seg 1/1 to reverse of: AC016033 from: 1 to: 117693
226 ValThrSerLeuGluAsnLeuArgAla 234
|||||
5726 GTCACCTCTCTGGAGATCTCAGGGCT 5700

```


OM of: US-09-701-711-2 to: Issued_Patents_NA:* out_format : pfs
 Date: Sep 19, 2002 4:43 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+p2n.model -DEV=xlp
 -O/cgn2_1/USPTO.scool/US09701711/runat_17092002_141431_28906/app_query.fasta_1.882
 -DB=Issued_Patents_NA -QFMF=fastap -SUFFIX=sepl7oli.rni
 -GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000
 -XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000
 -YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=oligo -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
 -THR_SCORE=quality -THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFMF=pfs
 -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09701711_@CGN1_1_104 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-701-711-2
 Query length: 813
 Database: Issued_Patents_NA:*
 Database sequences: 383533
 Database length: 122816752
 Search time (sec): 87.960000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
 WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	zScore	Escore Len	Documentation
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-936-165A-220	-	9.00	140.06	4.09
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-068-880-1	+	8.00	124.71	29.29
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-068-880-14	+	8.00	120.67	49.20
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-126-109-11	+	8.00	110.16	189.24
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-314-917-1	+	8.00	110.01	193.02
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-265-046-1	+	8.00	110.01	193.02
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-522-1	+	8.00	110.01	193.02
/cgn2_6/ptodata/2/ina/PTCTUS_COMB.seq:PCT-US93-11401-1	+	8.00	110.01	193.02
/cgn2_6/ptodata/2/ina/PTCTUS_COMB.seq:PCT-US95-07849-1	+	8.00	110.01	193.02
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-378-313-20	-	8.00	104.04	415.04
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2	-	8.00	56.82	1.6e+05
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1	-	8.00	56.80	1.6e+05
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-750-080A-6	+	7.00	118.17	67.75
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-651-472-6	+	7.00	118.17	67.75
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-358-928-6	+	7.00	118.17	67.75
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-611-757-96	-	7.00	113.58	122.17
/cgn2_6/ptodata/2/ina/PTCTUS_COMB.seq:PCT-US95-05980-96	-	7.00	113.58	122.17
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-143-576-8	-	7.00	113.29	126.72
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-13	-	7.00	113.08	130.14
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-11	-	7.00	113.05	130.70
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-10	-	7.00	113.02	131.27
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-5	-	7.00	112.95	132.41
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-840-683-4	+	7.00	112.91	132.98
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-555-722-4	+	7.00	112.91	132.98
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-384-301-4	+	7.00	112.91	132.98
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-12	-	7.00	112.91	132.98
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-485-657A-21	-	7.00	112.59	138.66
/cgn2_6/ptodata/2/ina/PTCTUS_COMB.seq:PCT-US95-02303-20	-	7.00	112.59	138.66
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-818-112-46	-	7.00	110.20	188.47
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-818-111-46	-	7.00	110.20	188.47
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-056-556-46	-	7.00	110.20	188.47
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-467-420A-5	-	7.00	110.03	192.42
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-110A-5	-	7.00	110.03	192.42
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-667-769A-5	-	7.00	110.03	192.42
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-940-371-5	-	7.00	110.03	192.42
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-637-647-5	-	7.00	110.03	192.42
/cgn2_6/ptodata/2/ina/PTCTUS_COMB.seq:PCT-US95-17082A-5	-	7.00	110.03	192.42
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-441-971-10	+	7.00	109.90	195.81

/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-221-653-10 + 7.00 109.90 195.81
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-442-144A-10 + 7.00 109.90 195.81
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-441-970-10 + 7.00 109.90 195.81
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-818-112-116 - 7.00 108.73 227.33
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-818-111-111 - 7.00 108.73 227.33
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-056-556-116 - 7.00 108.73 227.33
 /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-228-986-33 + 7.00 108.64 230.14

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-936-165A-220

seq_documentation_block:

; Sequence 220, Application US/08936165A
 ; Patent No. 6348582
 ; GENERAL INFORMATION:
 ; APPLICANT: Black, Michael
 ; APPLICANT: Burnham, Martin
 ; APPLICANT: Hodgson, John
 ; APPLICANT: Knowles, David
 ; APPLICANT: Lonetto, Michael
 ; APPLICANT: Nicholas, Richard
 ; APPLICANT: Pratt, Julie
 ; APPLICANT: Reichard, Richard
 ; APPLICANT: Rosenberg, Martin
 ; APPLICANT: Ward, Judith
 ; TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
 ; TITLE OF INVENTION: Polypeptides and Their Uses
 ; NUMBER OF SEQUENCES: 534
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SmithKline Beecham Corporation
 ; STREET: 709 Swedeland Road
 ; CITY: King of Prussia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/936,165A
 FILING DATE: 24-SEP-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/027,032
 FILING DATE: 24-SEP-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Gimmi, Edward R
 REGISTRATION NUMBER: 38,891
 REFERENCE/DOCKET NUMBER: P50549
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 610-270-4478
 TELEFAX: 610-270-5090
 TELEX:
 INFORMATION FOR SEQ ID NO: 220:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1000 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Genomic DNA
 US-08-936-165A-220

alignment_scores:

Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x US-08-936-165A-220/rev ..

Align seg 1/1 to reverse of: US-08-936-165A-220 from: 1 to: 1000

```
717 ValArgproValIlePheIleGluGly 725
|||||
311 GTGGACCGCTCATTTTATTGAAGGA 285

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-068-880-1

seq_documentation_block:
; Sequence 1, Application US/09068880B
; Patent No. 6203982
; GENERAL INFORMATION:
; APPLICANT: Nunokawa, Youichi
; APPLICANT: Oikawa, Shinzo
; APPLICANT: Tanaka, Shoji
; TITLE OF INVENTION: Method for Screening Compounds
; TITLE OF INVENTION: Regulating the Expression of Human-Inducible Nitric Oxide
; TITLE OF INVENTION: Synthase
; FILE REFERENCE: SHIM-001
; CURRENT APPLICATION NUMBER: US/09/068,880B
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: PCT/JP97/03303
; EARLIER FILING DATE: 1997-09-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-068-880-1

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-068-880-1 ..
Align seg 1/1 to: US-09-068-880-1 from: 1 to: 604

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
107 GAGGGCTACAGGAGGGTTAAAG 130

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-068-880-14

seq_documentation_block:
; Sequence 14, Application US/09068880B
; Patent No. 6203982
; GENERAL INFORMATION:
; APPLICANT: Nunokawa, Youichi
; APPLICANT: Oikawa, Shinzo
; APPLICANT: Tanaka, Shoji
; TITLE OF INVENTION: Method for Screening Compounds
; TITLE OF INVENTION: Regulating the Expression of Human-Inducible Nitric Oxide
; TITLE OF INVENTION: Synthase
; FILE REFERENCE: SHIM-001
; CURRENT APPLICATION NUMBER: US/09/068,880B
; CURRENT FILING DATE: 1998-09-02
; EARLIER APPLICATION NUMBER: PCT/JP97/03303
; EARLIER FILING DATE: 1997-09-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-068-880-14

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

717 ValArgproValIlePheIleGluGly 725
|||||
311 GTGGACCGCTCATTTTATTGAAGGA 285

Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-068-880-14 ..
Align seg 1/1 to: US-09-068-880-14 from: 1 to: 1026

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
2 GAGGGCTACAGGAGGGTTAAAG 25

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-126-109-11

seq_documentation_block:
; Sequence 11, Application US/09126109
; Patent No. 6171856
; GENERAL INFORMATION:
; APPLICANT: Thigpen, Anice
; APPLICANT: Hohmeier, Hans-Ewald
; APPLICANT: Newgard, Christopher B.
; APPLICANT: Unger, Roger H.
; APPLICANT: Shimabukuro, Michio
; APPLICANT: Chen, Guaxun
; APPLICANT: Rhodes, Christopher J.
; APPLICANT: Hugl, Sigrun R.
; APPLICANT: Cousin, Sharon
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TITLE OF INVENTION: TO NO-MEDIATED CYTOTOXICITY
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/126,109
; FILING DATE: 30-JUL-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,092
; FILING DATE: 30-JUL-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US Unknown
; FILING DATE: 03-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: McMillian, Nabeela R.
; REGISTRATION NUMBER: P-43,363
; REFERENCE/DOCKET NUMBER: UTSD:560
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4062 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-126-109-11

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-701-711-2 x US-09-126-109-11 ..
Align seg 1/1 to: US-09-126-109-11 from: 1 to: 4062

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
3566 GAGGGCCTACAGAGGGGTTAAAG 3589

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-314-917-1

seq_documentation_block:
; Sequence 1, Application US/08314917
; Patent No. 5468630
; GENERAL INFORMATION:
; APPLICANT: Billiar, Timothy R.
; APPLICANT: Nussler, Andreas K.
; APPLICANT: Geller, David A.
; APPLICANT: Simmons, Richard L.
; TITLE OF INVENTION: cDNA Clone for Human Inducible Nitric
; OXIDE SYNTHASE AND PROCESS FOR PREPARING SAME
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold B. Silverman
; STREET: Eckert Seamans Cherin & Mellott
; CITY: Pittsburgh
; STATE: PA
; COUNTRY: USA
; ZIP: 15219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/314,917
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/981,344
; FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Silverman, Arnold B.
; REGISTRATION NUMBER: 22,614
; REFERENCE/DOCKET NUMBER: 116972
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (412) 566-6000
; TELEFAX: (412) 566-6099
; TELEX: 866172
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide
; DESCRIPTION: Synthase cDNA Clone
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zap II cDNA
; CLONE: PHINOS
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: CDS

; LOCATION: 207..3668
; IDENTIFICATION METHOD: Experiment
US-08-314-917-1

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-314-917-1 ..
Align seg 1/1 to: US-08-314-917-1 from: 1 to: 4145

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
3667 GAGGGCCTACAGAGGGGTTAAAG 3690

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-265-046-1

seq_documentation_block:
; Sequence 1, Application US/08265046
; Patent No. 5658565
; GENERAL INFORMATION:
; APPLICANT: Timothy R. Billiar
; APPLICANT: Edith Tzeng
; APPLICANT: Andreas K. Nussler
; APPLICANT: David A. Geller
; APPLICANT: Richard L. Simmons
; TITLE OF INVENTION: Inducible Nitric Oxide Synthase
; GENE FOR TREATMENT OF DISEASE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lewis F. Gould, Jr.
; ADDRESS: Eckert Seamans Cherin & Mellott
; STREET: 1700 Market Street, Suite 3232
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/265-046
; FILING DATE: 24-JUN-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Gould, Lewis F. Jr.
; REGISTRATION NUMBER: 25,057
; REFERENCE/DOCKET NUMBER: 119130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 575-6020
; TELEFAX: (215) 575-6015
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide
; DESCRIPTION: Synthase cDNA Clone
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zap II cDNA
```

```

; CLONE: PHINOS
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 207..3668
; IDENTIFICATION METHOD: Experiment
US-08-265-046-1

```

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-09-701-711-2 x US-08-265-046-1 ..
Align seg 1/1 to: US-08-265-046-1 from: 1 to: 4145

```

115 GluGlyLeuGlnGluGlyLeuLys 122

```

|||||
3667 GAGGCGCTACAGGAGGGTTAAAG 3690

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-465-522-1

seq_documentation_block:

```

; Sequence 1, Application US/08465522
; Patent No. 5882908
; GENERAL INFORMATION:
; APPLICANT: Billiar, Timothy R.
; APPLICANT: Nussler, Andreas K.
; APPLICANT: Geller, David A.
; APPLICANT: Simmons, Richard L.
; TITLE OF INVENTION: cDNA Clone for Human Inducible Nitric
; TITLE OF INVENTION: Oxide Synthase And Process for Preparing Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lewis F. Gould, Jr.
; ADDRESSEE: Eckert Seamans, Cherin & Mellott
; STREET: 1700 Market St. Suite 3232
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,522
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gould, Jr., Lewis F.
; REGISTRATION NUMBER: 25,057
; REFERENCE/DOCKET NUMBER: 116972-6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 575-6020
; TELEFAX: (215) 575-6015
; TELEX:

```

INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide
; DESCRIPTION: Synthase cDNA Clone

```

```

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda Zap II cDNA
; CLONE: PHINOS
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 207..3668
; IDENTIFICATION METHOD: Experiment
US-08-465-522-1

```

```

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

```

US-09-701-711-2 x US-08-465-522-1 ..
Align seg 1/1 to: US-08-465-522-1 from: 1 to: 4145

```

115 GluGlyLeuGlnGluGlyLeuLys 122

```

|||||
3667 GAGGCGCTACAGGAGGGTTAAAG 3690

```

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US93-11401-1

seq_documentation_block:

```

; Sequence 1, Application PCTUS9311401
; GENERAL INFORMATION:
; APPLICANT: Billiar, Timothy R.
; APPLICANT: Nussler, Andreas K.
; APPLICANT: Geller, David A.
; APPLICANT: Simmons, Richard L.
; TITLE OF INVENTION: cDNA Clone for Human Inducible Nitric
; TITLE OF INVENTION: Oxide Synthase And Process for Preparing Same
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold B. Silverman
; ADDRESSEE: Eckert Seamans Cherin & Mellott
; STREET: 600 Grant Street, 42nd Floor
; CITY: Pittsburgh
; STATE: PA
; COUNTRY: USA
; ZIP: 15219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/11401
; FILING DATE: 25-NOV-1992
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/981,344
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Silverman, Arnold B.
; REGISTRATION NUMBER: 22,614
; REFERENCE/DOCKET NUMBER: 116972
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (412) 566-6000
; TELEFAX: (412) 566-6099
; TELEX: 866172
; INFORMATION FOR SEQ ID NO: 1:

```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zap II cdna
; CLONE: PHINOS
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 207..3668
; IDENTIFICATION METHOD: Experiment
; PCT-US93-11401-1

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x PCT-US93-11401-1 ..
Align seg 1/1 to: PCT-US93-11401-1 from: 1 to: 4145

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
3667 GAGGGCCTACAGGAGGGTTAAAG 3690

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-07849-1

seq_documentation_block:
; Sequence 1, Application PC/TUS9507849
; GENERAL INFORMATION:
; APPLICANT: University of Pittsburgh of the Commonwealth System of Higher
; APPLICANT: Education
; TITLE OF INVENTION: Inducible Nitric Oxide Synthase
; TITLE OF INVENTION: Gene for Treatment of Disease
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lewis F. Gould, Jr.
; ADDRESSEE: Eckert Seamans Cherin & Mellott
; STREET: 1700 Market Street, Suite 3232
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07849
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gould, Lewis F. Jr.
; REGISTRATION NUMBER: 25,057
; REFERENCE/DOCKET NUMBER: 119130-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 575-6020
; TELEFAX: (215) 575-6015
; TELEX:
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4145 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; DESCRIPTION: Human Hepatocyte Inducible Nitric Oxide
; DESCRIPTION: Synthase cdna Clone
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; TISSUE TYPE: Induced Human Hepatocyte RNA
; IMMEDIATE SOURCE:
; LIBRARY: Lambda zap II cdna
; CLONE: PHINOS
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: unknown
; MAP POSITION: unknown
; UNITS: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 207..3668
; IDENTIFICATION METHOD: Experiment
; PCT-US95-07849-1

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x PCT-US95-07849-1 ..
Align seg 1/1 to: PCT-US95-07849-1 from: 1 to: 4145

115 GluGlyLeuGlnGluGlyLeuLys 122
|||||
3667 GAGGGCCTACAGGAGGGTTAAAG 3690

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-378-313-20

seq_documentation_block:
; Sequence 20, Application US/08378313
; Patent No. 6207881
; GENERAL INFORMATION:
; APPLICANT: THEOLOGIS, ATHANASIOS
; APPLICANT: SATO, TAKAHIDO
; TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH
; TITLE OF INVENTION: GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378.313
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/862,493
; FILING DATE: 02-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: MURASHIGE, KATE H.
```

```
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 29190-20002.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 856-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2704..2880, 2968..3099, 3183..3344, 3810
; LOCATION: ..4376, 4463..4903)
; US-08-378-313-20
```

```
alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-701-711-2 x US-08-378-313-20/rev ..
```

```
Align seg 1/1 to reverse of: US-08-378-313-20 from: 1 to: 9060
```

```
500 ThrLysTyAspAsnLysAsnIle 507
|||||
5398 ACAAAATATGACACAAAAACATC 5375
```

```
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-2
```

```
seq_documentation_block:
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
```

```
alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-701-711-2 x US-09-103-840A-2/rev ..
```

```
Align seg 1/1 to reverse of: US-09-103-840A-2 from: 1 to: 4403765
```

```
685 ThrThrLeuGlyGluValValGly 692
|||||
```

```
3909845ACCACGCTGGGGAGGTGTCGGA 3909822
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-840A-1
```

```
seq_documentation_block:
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
```

```
alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-701-711-2 x US-09-103-840A-1/rev ..
```

```
Align seg 1/1 to reverse of: US-09-103-840A-1 from: 1 to: 4411529
```

```
685 ThrThrLeuGlyGluValValGly 692
|||||
3916067ACCACGCTGGGGAGGTGTCGGA 3916044
```

```
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-750-080A-6
```

```
seq_documentation_block:
; Sequence 6, Application US/07750080A
; Patent No. 5445953
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF A MODIFIED
; TITLE OF INVENTION: EUKARYOTIC CYTOPLASMIC DNA VIRUS GENOME
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/750,080A
; FILING DATE: 19910826
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/106 IMMU
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: PA0 (Fig. 4.3)
US-07-750-080A-6

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-07-750-080A-6 ..

Align seg 1/1 to: US-07-750-080A-6 from: 1 to: 115

409 ArgProValProAsnSerPro 415
|||||
90 AGGCCAGTACCCCAATTCGCC 110

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-651-472-6

seq_documentation_block:
Sequence 6, Application US/08651472
Patent No. 6103244

GENERAL INFORMATION:
APPLICANT: DORNER, Friedrich
APPLICANT: SCHEIFLINGER, Friedrich
APPLICANT: FALKNER, Falko Gunter
APPLICANT: PFEIDERER, Michael
TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/651,472
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,738
FILING DATE: 20-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/750,080
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/166/IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399

TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 115 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Other nucleic acid;
DESCRIPTION: Synthetic DNA oligonucleotide
IMMEDIATE SOURCE:
CLONE: PA0
US-08-651-472-6

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-651-472-6 ..

Align seg 1/1 to: US-08-651-472-6 from: 1 to: 115

409 ArgProValProAsnSerPro 415
|||||
90 AGGCCAGTACCCCAATTCGCC 110

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-358-928-6

seq_documentation_block:
Sequence 6, Application US/08358928
Patent No. 6265183

GENERAL INFORMATION:
APPLICANT: DORNER, Friedrich
APPLICANT: SCHEIFLINGER, Friedrich
APPLICANT: FALKNER, Falko Gunter
APPLICANT: PFEIDERER, Michael
TITLE OF INVENTION: DIRECT MOLECULAR CLONING OF CHIMERIC
TITLE OF INVENTION: VIRUSES CONTAINING HUMAN IMMUNODEFICIENCY VIRUS TYPE 1
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,928
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/914,738
FILING DATE: 20-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/750,080
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/166/IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 6:

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid;
; DESCRIPTION: Synthetic DNA oligonucleotide
; IMMEDIATE SOURCE:
; CLONE: pA0
; US-08-358-928-6

alignment_scores:
  Quality: 7.00      Length:
  Ratio: 1.000      Gaps:
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-358-928-6 ..
Align seg 1/1 to: US-08-358-928-6 from: 1 to: 115
409 ArgProValProAsnSerPro 415
|||||
90 AGGCAGTACCCATTCGCC 110

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-611-757-96.

seq_documentation_block:
; Sequence 96, Application US/08611757
; Patent No. 5859230
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Reyes, Gregory R.
; APPLICANT: Wages, John
; APPLICANT: Zhang-keck, Zhen-yang
; APPLICANT: Young, Lavonne
; TITLE OF INVENTION: No. 5859230-A/No. 5859230-B/No. 5859230-C/No. 5859230-D/No. 58
; TITLE OF INVENTION: Agents and Molecular Cloning Thereof
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,757
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/246,985
; FILING DATE: 20-MAY-1994
; APPLICATION NUMBER: US 025,396
; FILING DATE: 24-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/922,493
; FILING DATE: 30-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 115 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SCH Clone SU6-2
```

```
; LENGTH: 210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SCH Clone SU6-2
; US-08-611-757-96

alignment_scores:
  Quality: 7.00      Length:
  Ratio: 1.000      Gaps:
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-611-757-96/rev ..
Align seg 1/1 to reverse of: US-08-611-757-96 from: 1 to: 210
787 SerIleSerTyrAlaLysPro 793
|||||
91 TCGATTTCCTACGCCAAACCC 71

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-05980-96

seq_documentation_block:
; Sequence 96, Application PC/TUS9505980
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Non-A/Non-B/Non-C/Non-D/Non-E Hepatitis
; TITLE OF INVENTION: Agents and Molecular Cloning Thereof
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05980
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,986
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.49
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SCH Clone SU6-2
```


PCT-US95-05980-96

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x PCT-US95-05980-96/rev ..

Align seg 1/1 to reverse of: PCT-US95-05980-96 from: 1 to: 210

787.SerIleSerTyrAlaLysPro 793
|||||
91 TCGATTTCCTAGCCCAACCC 71

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-143-576-8

seq_documentation_block:
; Sequence 8, Application US/08143576
; Patent No. 5643761

GENERAL INFORMATION:
; APPLICANT: Fisher, Paul B.
; APPLICANT: Jiang, Hongping
; TITLE OF INVENTION: METHOD FOR GENERATING A SUBSTRACTED
; TITLE OF INVENTION: CDNA LIBRARY AND USES OF THE GENERATED LIBRARY
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRES:
; ADDRESSEE: John P. White, c/o Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/143,576
; FILING DATE: 25-OCT-1993
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 43563/JPW/AKC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-143-576-8

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-143-576-8/rev ..

Align seg 1/1 to reverse of: US-08-143-576-8 from: 1 to: 218

472 ArgGluValTyrSerLeuGly 478
|||||
216 CGAGAGGTTTACAGCTAGGT 196

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-13

seq_documentation_block:
; Sequence 13, Application US/09157270
; Patent No. 6306401

GENERAL INFORMATION:
; APPLICANT: Johnson, Eric S.
; APPLICANT: Pham, Thuy D.
; TITLE OF INVENTION: Viral Detection System
; FILE REFERENCE: D6161
; CURRENT APPLICATION NUMBER: US/09/157,270
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/061,287
; EARLIER FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 13
; LENGTH: 224
; TYPE: DNA
; ORGANISM: unknown

FEATURE:
; OTHER INFORMATION: RT-PCR product from egg albumin of commercial chicken from
; OTHER INFORMATION: randomly chosen grocery store # 205.
US-09-157-270-13

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-157-270-13/rev ..

Align seg 1/1 to reverse of: US-09-157-270-13 from: 1 to: 224

289 GluLeuGluAlaLeuLys 295
|||||
189 GAGTTAGAACGCGTGTAAAG 169

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-11

seq_documentation_block:
; Sequence 11, Application US/09157270
; Patent No. 6306401

GENERAL INFORMATION:
; APPLICANT: Johnson, Eric S.
; APPLICANT: Pham, Thuy D.
; TITLE OF INVENTION: Viral Detection System
; FILE REFERENCE: D6161
; CURRENT APPLICATION NUMBER: US/09/157,270
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/061,287
; EARLIER FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 11
; LENGTH: 225
; TYPE: DNA
; ORGANISM: Single Comb White Leghorn chicken

FEATURE:
; OTHER INFORMATION: RT-PCR product from egg albumin of stock 70 chicken.
US-09-157-270-11

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-157-270-11/rev ..
Align seg 1/1 to reverse of: US-09-157-270-11 from: 1 to: 225

289 GluLeuGluAlaLeuLeuLys 295
|||||
191 GAGTTAGAGCGCTGTTAAAG 171

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-10

seq_documentation_block:
; Sequence 10, Application US/09157270
; Patent No. 6306401

; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric S.
; APPLICANT: Pham, Thuy D.
; TITLE OF INVENTION: Viral Detection System
; FILE REFERENCE: D6161
; CURRENT APPLICATION NUMBER: US/09/157,270
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/061,287
; EARLIER FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 10
; LENGTH: 226
; TYPE: DNA
; ORGANISM: Single Comb White Leghorn chicken
; FEATURE:
; OTHER INFORMATION: RT-PCR product from egg albumin of stock 6F chicken.
US-09-157-270-10

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-157-270-10/rev ..

Align seg 1/1 to reverse of: US-09-157-270-10 from: 1 to: 226

289 GluLeuGluAlaLeuLeuLys 295
|||||
191 GAGTTAGAGCGCTGTTAAAG 171

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-5

seq_documentation_block:
; Sequence 5, Application US/09157270
; Patent No. 6306401

; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric S.
; APPLICANT: Pham, Thuy D.
; TITLE OF INVENTION: Viral Detection System
; FILE REFERENCE: D6161
; CURRENT APPLICATION NUMBER: US/09/157,270
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/061,287
; EARLIER FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 5
; LENGTH: 228
; TYPE: DNA
; ORGANISM: Avian leukosis/sarcoma virus of the family Retroviridae
; FEATURE:
; OTHER INFORMATION: Subgroup E of the avian leukosis/sarcoma virus.
US-09-157-270-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x US-09-157-270-5/rev ..

Align seg 1/1 to reverse of: US-09-157-270-5 from: 1 to: 228

289 GluLeuGluAlaLeuLeuLys 295
|||||
194 GAGTTAGAGCGCTGTTAAAG 174

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-840-683-4

seq_documentation_block:
; Sequence 4, Application US/08840683
; Patent No. 5821051

; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: E6-BINDING PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,683
; FILING DATE: 29-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/273,059
; FILING DATE: 08-JUL-1994

; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-840-683-4

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-840-683-4 ..

Align seg 1/1 to: US-08-840-683-4 from: 1 to: 229

52 ValLeuProPheArgLeuGly 58
|||||
58 GTTTTACCATTAGCTGGCT 78

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-555-722-4

```
seq_documentation_block:
; Sequence 4, Application US/08555722
; Patent No. 5989804
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: EG-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,722
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-555-722-4

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-555-722-4 ..
Align seg 1/1 to: US-08-555-722-4 from: 1 to: 229

52 ValLeuPropheArgLeuGly 58
|||||
58 GTTTACCATTTAGCTGGGT 78

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-384-301-4

seq_documentation_block:
; Sequence 4, Application US/09384301
; Patent No. 6296853
; GENERAL INFORMATION:
; APPLICANT: Androphy, Elliot J.
; APPLICANT: Chen, Jason J.
; TITLE OF INVENTION: EG-BINDING PROTEINS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
```

```
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/384,301
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/555,722
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Myers, Louis
; REGISTRATION NUMBER: 35,965
; REFERENCE/DOCKET NUMBER: NEP-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 229 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-09-384-301-4

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-09-384-301-4 ..
Align seg 1/1 to: US-09-384-301-4 from: 1 to: 229

52 ValLeuPropheArgLeuGly 58
|||||
58 GTTTACCATTTAGCTGGGT 78

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-157-270-12

seq_documentation_block:
; Sequence 12, Application US/09157270
; Patent No. 6306401
; GENERAL INFORMATION:
; APPLICANT: Johnson, Eric S.
; APPLICANT: Pham, Thuy D.
; TITLE OF INVENTION: Viral Detection System
; FILE REFERENCE: D6161
; CURRENT APPLICATION NUMBER: US/09/157,270
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: US 60/061,287
; EARLIER FILING DATE: 1997-10-07
; NUMBER OF SEQ ID NOS: 26
; SEQ ID NO 12
; LENGTH: 229
; TYPE: DNA
; ORGANISM: Single Comb White Leghorn chicken
; FEATURE:
; OTHER INFORMATION: RT-PCR product from egg albumin of stock 100 chicken.
US-09-157-270-12

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
```

```
US-09-701-711-2 x US-09-157-270-12/rev ..
Align seg 1/1 to reverse of: US-09-157-270-12 from: 1 to: 229
289 GluLeuGluAlaLeuLeuLys 295
|||||
194 GAGTTAGAAGCGCTGTTAAG 174

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-485-657A-21
seq_documentation_block:
; Sequence 21, Application US/08485657A
; Patent No. 5942389
; GENERAL INFORMATION:
; APPLICANT: Kirschling, Deborah J
; APPLICANT: Gudkov, Andrei
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Genes and Genetic Elements Associated
; TITLE OF INVENTION: with Sensitivity to Cisplatin
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/485,657A
; APPLICATION NUMBER: 07-JUN-1995
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5942389man, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-N
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ANTI-SENSE: YES
US-08-485-657A-21

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-485-657A-21/rev ..
Align seg 1/1 to reverse of: US-08-485-657A-21 from: 1 to: 239
289 GluLeuGluAlaLeuLeuLys 295
|||||
153 GAACGGAGGCGCTCTGAAA 133

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-02303-20
seq_documentation_block:
; Sequence 20, Application PC/TUS9502303
```

```
; GENERAL INFORMATION:
; APPLICANT: Genes and Genetic Elements Associated
; TITLE OF INVENTION: with Sensitivity to Cisplatin
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION NUMBER: PCT/US95/02303
; APPLICATION NUMBER: PCT/US95/02303
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ANTI-SENSE: YES
; PCT-US95-02303-20

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x PCT-US95-02303-20/rev ..
Align seg 1/1 to reverse of: PCT-US95-02303-20 from: 1 to: 239
289 GluLeuGluAlaLeuLeuLys 295
|||||
153 GAACGGAGGCGCTCTGAAA 133

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-818-112-46
seq_documentation_block:
; Sequence 46, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-46

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x US-08-818-112-46/rev ..

Align seg 1/1 to reverse of: US-08-818-112-46 from: 1 to: 327

259 ArgillePheValGlulIleSer 265
|||||
226 CGAATATTTCGTCGAGATCTCG 206

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-818-111-46

seq_documentation_block:

; Sequence 46, Application US/08818111
; Patent No. 6338852

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 46:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 327 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-818-111-46

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x US-08-818-111-46/rev ..

Align seg 1/1 to reverse of: US-08-818-111-46 from: 1 to: 327

259 ArgillePheValGlulIleSer 265
|||||
226 CGAATATTTCGTCGAGATCTCG 206

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-056-556-46

seq_documentation_block:

; Sequence 46, Application US/09056556
; Patent No. 6350456

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 46:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 327 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-09-056-556-46

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x US-09-056-556-46/rev ..

Align seg 1/1 to reverse of: US-09-056-556-46 from: 1 to: 327

259 ArgillePheValGlulIleSer 265
|||||
226 CGAATATTTCGTCGAGATCTCG 206

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-467-420A-5

TRE

seq_documentation_block:
; Sequence 5, Application US/08467420A
; Patent No. 5683892
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: Holmes, Stephen D.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TITLE OF INVENTION: Treatment of IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESSEE: Intellectual Property
; STREET: P. O. Box 1539-UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,420A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363131
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270-5024
; TELEFAX: 610 270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; OTHER INFORMATION: Kabat position 24"
US-08-467-420A-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-467-420A-5/rev ..

Align seg 1/1 to reverse of: US-08-467-420A-5 from: 1 to: 334

669 LeuGlyProArgSerGlnAla 675
|||||
311 CTGGCCCCAGAGTCAGGCC 291

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-470-110A-5

seq_documentation_block:
; Sequence 5, Application US/08470110A
; Patent No. 5693323
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: Holmes, Stephen D.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TITLE OF INVENTION: Treatment of IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESSEE: Intellectual Property
; STREET: P. O. Box 1539-UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,110A
; FILING DATE:
; CLASSIFICATION: 426
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363131
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270-5024
; TELEFAX: 610 270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; OTHER INFORMATION: Kabat position 24"
US-08-470-110A-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x US-08-470-110A-5/rev ..

Align seg 1/1 to reverse of: US-08-470-110A-5 from: 1 to: 334

669 LeuGlyProArgSerGlnAla 675
|||||
311 CTGGCCCCAGAGTCAGGCC 291

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-667-769A-5

```
seq_documentation_block:
; Sequence 5, Application US/08667769A
; Patent No. 5783184
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: Holmes, Stephen D.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TITLE OF INVENTION: Treatment of IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; STREET: P.O. Box 1539-UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/667,769A
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17082
; FILING DATE: 22-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470110
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467420
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363131
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50503
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5024
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; US-08-667-769A-5
; Kabat position 24"
```

```
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x US-08-667-769A-5/rev ..
```

```
Align seq 1/1 to reverse of: US-08-667-769A-5 from: 1 to: 334
669 LeuGlyProArgSerGlnAla 675
|||||
311 CTTGGCCCCCAGAAAGTCAAGCC 291

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-940-371-5

seq_documentation_block:
; Sequence 5, Application US/08940371
; Patent No. 5851525
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: Holmes, Stephen D.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TITLE OF INVENTION: Treatment of IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESS: Intellectual Property
; STREET: P. O. Box 1539-UW2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/940,371
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,110
; FILING DATE:
; APPLICATION NUMBER: US 08/363131
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270-5024
; TELEFAX: 610 270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; US-08-940-371-5
; Kabat position 24"
```

```
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
```

US-09-701-711-2 x US-08-940-371-5/rev ..

Align seg 1/1 to reverse of: US-08-940-371-5 from: 1 to: 334

669 LeuGlyProArgSerGlnAla 675
|||||
311 CTGGCCCCAGAGTCAAGCC 291

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-637-647-5

seq documentation block:
; Sequence 5, Application US/08637647
; Patent No. 6129913
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TITLE OF INVENTION: Treatment of IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; ADDRESSEE: Intellectual Property
; STREET: P. O. Box 1539-UM2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,647
; FILING DATE: 19-JUN-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,131
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610 270-5024
; TELEFAX: 610 270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; OTHER INFORMATION: Kabat position 24"
US-08-637-647-5

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:

US-09-701-711-2 x US-08-637-647-5/rev ..

Align seg 1/1 to reverse of: US-08-637-647-5 from: 1 to: 334

669 LeuGlyProArgSerGlnAla 675
|||||
311 CTGGCCCCAGAGTCAAGCC 291

seq_name: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:PCT-US95-17082A-5

seq documentation block:
; Sequence 5, Application PC/TUS9517082A
; GENERAL INFORMATION:
; APPLICANT: Ames, Robert S.
; APPLICANT: Appelbaum, Edward R.
; APPLICANT: Chaiken, Irwin M.
; APPLICANT: Cook, Richard M.
; APPLICANT: Gross, Mitchell S.
; APPLICANT: Holmes, Stephen D.
; APPLICANT: McMillan, Lynette J.
; APPLICANT: Theisen, Timothy W.
; TITLE OF INVENTION: Recombinant IL5 Antagonists Useful in
; TITLE OF INVENTION: Treatment of IL5 Mediated Disorders
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corp./Corporate
; STREET: P.O. Box 1539-UM2220
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/17082A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/470110
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/467420
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/363131
; FILING DATE: 23-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sutton, Jeffrey A.
; REGISTRATION NUMBER: 34,028
; REFERENCE/DOCKET NUMBER: P50282-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5024
; TELEFAX: 610-270-5090
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 334 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..334
; OTHER INFORMATION: /note= "First base corresponds to
; OTHER INFORMATION: Kabat position 24"
PCT-US95-17082A-5

alignment_scores:
Quality: 7.00 Length: 7

Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x PCT-US95-17082A-5/rev ..
Align seg 1/1 to reverse of: PCT-US95-17082A-5 from: 1 to: 334
seq_name: /cgn2_6/ptodata/2/ins/6A_COMB.seq:US-08-441-971-10
669 LeuGlyProArgSerGlnAla 675
311 CTGTGCCCCGAGAGTCAGCC 291
seq_documentation_block:
; Sequence 10, Application US/08441971
; Patent No. 6071693
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441,971
; FILING DATE: 16-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,653
; FILING DATE:
; FILING DATE: 07/697,326
; FILING DATE:
; FILING DATE: 8 May 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janluk, Anthony J.
; REGISTRATION NUMBER: 29,809
; REFERENCE/DOCKET NUMBER: C0772/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 720-3500
; TELEFAX: (617) 720-2441
; TELEX: EZEKIEL
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5gh6
US-08-441-971-10
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x US-08-441-971-10 ..

Align seg 1/1 to: US-08-441-971-10 from: 1 to: 340
669 LeuGlyProArgSerGlnAla 675
58 CTGTGCCCCCGAAGCCAGGCA 78
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-221-653-10
seq_documentation_block:
; Sequence 10, Application US/08221653
; Patent No. 6190864
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS Version 3.3
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,653
; FILING DATE:
; FILING DATE: 07/697,326
; FILING DATE:
; FILING DATE: 8 May 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Janluk, Anthony J.
; REGISTRATION NUMBER: 29,809
; REFERENCE/DOCKET NUMBER: C0772/7000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 720-3500
; TELEFAX: (617) 720-2441
; TELEX: EZEKIEL
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5gh6
US-08-221-653-10
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-701-711-2 x US-08-221-653-10 ..
Align seg 1/1 to: US-08-221-653-10 from: 1 to: 340
669 LeuGlyProArgSerGlnAla 675
58 CTGTGCCCCCGAAGCCAGGCA 78
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-442-144A-10

```
seq_documentation_block:
; Sequence 10, Application US/08442144A
; Patent No. 6214583
; GENERAL INFORMATION:
; APPLICANT: Tai-An Cha
; APPLICANT: Eileen Beall
; APPLICANT: Bruce Irvine
; APPLICANT: Janice Kolberg
; APPLICANT: Michael S. Urdea
; TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
; NUMBER OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 Inch
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows NT
; SOFTWARE: Microsoft Word 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/442,144A
; FILING DATE: MAY 16, 1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/221,653
; FILING DATE: APRIL 1, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Doreen Yanko Trujillo
; REGISTRATION NUMBER: 35,719
; REFERENCE/DOCKET NUMBER: CHIR-0121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; TELEX:
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 340 Nucleotides
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: ns5gh6
; US-08-442-144A-10
```

```
alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
```

```
US-09-701-711-2 x US-08-442-144A-10 ..
```

```
Align seg 1/1 to: US-08-442-144A-10 from: 1 to: 340
```

```
669 LeuGlyProArgSerGlnAla 675
|||||
58 CTTGGCCCCCGAAGCCAGGCA 78
```

OM of: US-09-701-711-2 to: EST:* out_format : pfs

Date: Sep 19, 2002 3:44 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09701711/runat_17092002.141431_28874/app_query.fasta_1.882  
-DB=EST -QFWT=fastap -SUFFIX=sepl7oli.rst -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=60.000 -XGAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality  
-THR_MIN=1 -ALIGN=40 -MODE=LOCAL -OUTFWT=pfs -NORM=ext  
-HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09701711 -CGN1_1=5751 -NCPU=6 -ICPU=3 -LONGLOG  
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-701-711-2

Query length: 813

Database: EST:

Database sequences: 13736207

Database length: -1841457050

Search time (sec): 1940.450000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_est1:BM209113	-	9.00	155.68	20.04	180
gb_est1:AI528822	-	9.00	153.13	27.77	238
gb_est1:BM010871	-	9.00	151.69	33.43	279
gb_est1:BI723030	-	9.00	148.57	49.87	393
gb_est1:BE840995	-	9.00	148.45	50.61	398
gb_est1:AV625127	-	9.00	147.04	60.69	465
gb_gss:AAQ67844	+	9.00	146.90	61.75	472
gb_gss:AAQ837806	+	9.00	146.71	63.28	482
gb_gss:AAQ836677	-	9.00	145.25	76.34	566
gb_gss:BH025710	+	9.00	145.25	76.34	566
gb_gss:AAQ509377	-	9.00	144.84	80.44	592
gb_gss:AAQ290748	-	9.00	144.55	83.47	611
gb_est1:BI873688	+	9.00	144.32	86.02	627
gb_gss:AAZ391700	+	9.00	144.20	87.31	635
gb_est1:BI998945	+	9.00	144.00	89.56	649
gb_est1:BG177050	+	9.00	143.92	90.52	655
gb_gss:CNS04ADV	+	9.00	140.84	134.41	919
gb_est1:BG397547	-	9.00	137.18	214.93	1374
gb_est1:BG420056	-	9.00	135.81	256.17	1597
gb_est1:AW796076	-	8.00	140.24	145.08	116
gb_gss:AAZ286361	-	8.00	140.01	149.47	119
gb_gss:AAZ909005	+	8.00	138.80	174.67	136
gb_est1:BF983453	+	8.00	135.15	278.78	203
gb_est1:AV257342	-	8.00	135.02	283.60	206
gb_est1:BM128835	-	8.00	134.93	286.82	208
gb_est1:R70943	-	8.00	134.63	298.11	215
gb_est1:BF377641	+	8.00	134.54	301.35	217
gb_est1:BI493603	+	8.00	133.90	327.44	233
gb_est1:AA729757	-	8.00	133.48	345.56	244
gb_est1:BA817742	-	8.00	133.15	360.48	253
gb_est1:BB579201	+	8.00	132.97	368.81	258
gb_est1:AA247205	+	8.00	132.83	375.49	262
gb_est1:AI866213	+	8.00	132.62	385.55	268
gb_est1:AI583432	+	8.00	132.42	395.64	274
gb_est1:BI537706	+	8.00	132.32	400.70	277
gb_gss:AAZ375019	+	8.00	132.32	400.70	277
gb_est1:HB2436	-	8.00	131.81	427.84	293
gb_est1:AV205177	-	8.00	131.75	431.26	295

gb_est1:AU100452 + 8.00 131.60 439.80 300 ! AU100452 AU100452 Sugano Hom
gb_est1:AI528822 - 8.00 131.54 443.22 302 ! AI528822 ms15c10.x1 Strategic
gb_est1:BF428988 - 8.00 131.51 444.94 303 ! BF428988 WHE1712.G05.N102S.W
gb_est1:AA776501 + 8.00 131.42 450.08 306 ! AA776501 ah11f11.s1 Gessler
gb_est1:HG0311 + 8.00 131.42 450.08 306 ! HG0311 yr41e09.s1 Soares fet
gb_gss:BH596414 - 8.00 131.18 463.85 314 ! BH596414 B0GEA797F B0GE Brns
gb_est1:BG905044 - 8.00 131.01 474.21 320 ! BG905044 TaLr1136G07F TaLr1

seq_name: gb_est2:BM209113

seq_documentation_block:

LOCUS BM209113 180 bp mRNA linear EST 31-JAN-2002
C0640C12-3 NIA Mouse Trophoblast Stem Cell cDNA Library (Long) Mus
DEFINITION musculus cDNA clone C0640C12 3', mRNA sequence.

ACCESSION BM209113

VERSION BM209113.1 GI:17765496

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 180)

AUTHORS Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Luo,A.,
Tanaka,T., Kunath,T., Rossant,J. and Ko,M.S.H.

TITLE Systematic Analyses of NIA Mouse Trophoblast Stem Cell cDNA Library
(Long)

JOURNAL Unpublished (2001)

COMMENT Contact: Dawood B. Dudekula
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
Email: cdna@lgsun.grc.nia.nih.gov
Plate: C0640 row: C column: 12
Seq primer: -21M13 Forward
High quality sequence stop: 180
POLYA=Yes.

FEATURES

source

Location/Qualifiers

1..180

/organism="Mus musculus"

/strain="B5/EGFP transgenic ICR mice"

/db_xref="niaEST:C0640C12-3"

/db_xref="taxon:10090"

/clone="C0640C12"

/clone_lib="NIA Mouse Trophoblast Stem Cell cDNA Library
(Long)"

/tissue_type="Trophoblast stem cell"

/dev_stage="3.5-dpc"

/lab_host="DH10B"

/notes="Vector: pSPOT1 (Invitrogen); Site_1: SalI; Site_2:
NotI; Mouse cDNA project by the Laboratory of Genetics,
National Institute on Aging (NIA), Intramural Research
Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
a long-transcript enriched cDNA library (Ref. Genome Res.
11: 1553-1558 (2001)). [PMID: 11544199]. Total RNAs were
obtained from Dr. Janet Rossant and Tilo Kunath (Samuel
Lunenfeld Research Institute, Canada). Double-stranded
cDNAs were synthesized with an Oligo(dT) primer
(Invitrogen):
5'-pGACTAGTCTAGATCCGAGCGGCCCTTTTCTTTT-3' from
4 ug of total RNA, treated with T4 DNA polymerase, and
purified by ethanol-precipitation. The cDNAs were ligated
to Lona-linker LL-Sal4, purified by phenol/chloroform, and
separated from free linkers by Centricon 100. Then, the
cDNAs were amplified by long-range high fidelity PCR using
Ex Taq polymerase (Takara) with a primer Sal4-S. The
products were purified by phenol/chloroform and Centricon
100. The cDNAs were digested with SalI and NotI enzymes
and cloned into SalI/NotI site of pSPOT1 plasmid vector.
The DH10B E. coli host was transformed with the ligation
mixture by the standard chemical method. The average
insert size is about 2.6 kb. The library was constructed
by Yulan Piao (NIA)." 43 t

BASE COUNT 49 a 38 c 50 g

ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x BM209113/rev ..

Align seg 1/1 to reverse of: BM209113 from: 1 to: 180

705 LeuLeuProLeuProPheLysGlyAspTrrp 713

|||||
 53 CTGCCCTACCCCTCAAGGAGATTGG 27

seq_name: gb_estl:A1886635

seq_documentation_block:

LOCUS A1886635 238 bp mRNA linear EST 07-MAR-2000
 DEFINITION wm60h01.x1 NCI_CGAP_UT2 Homo sapiens cDNA clone IMAGE:2440369 3',
 mRNA sequence.

ACCESSION A1886635

VERSION A1886635.1 GI:5591799

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 238)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1944 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 217.

FEATURES

source

1..238
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2440369"
 /clone_lib="NCI_CGAP_UT2"
 /tissue_type="moderately-differentiated endometrial
 adenocarcinoma, 3 pooled tumors"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.85 kb. Life Technologies catalog #:
 11539-012"

BASE COUNT

ORIGIN

alignment_scores:

Quality: 9.00 Length: 84 t
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x A1886635 ..

Align seg 1/1 to: A1886635 from: 1 to: 238

703 LeuLeuProLeuProPheLysGly 711

|||||
 188 CTTATACCTCCACTTCCTTTAAAGGA 214

seq_name: gb_estl:BB010871

seq_documentation_block:

LOCUS BB010871 279 bp mRNA linear EST 22-JUN-2000
 DEFINITION BB010871 RIKEN full-length enriched, 10 day neonate skin Mus
 musculus cDNA clone 4732498B18 3', mRNA sequence.

ACCESSION BB010871

VERSION BB010871.1 GI:8131228

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 279)

AUTHORS

Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
 ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
 Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
 Iizawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
 , Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
 ,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya
 ,T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
 , Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
 ,M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Konno,H., et al.)
 Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsc.riken.go.jp,
 URL:<http://genome.gsc.riken.go.jp/>
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
 ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Thermotabilization and thermoactivation of thermostable enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
 ,Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (<http://genome.rtc.riken.go.jp>) for
 further details.

FEATURES

source

1..279
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="4732498B18"
 /clone_lib="RIKEN full-length enriched, 10 day neonate
 skin"
 /sex="mixed"
 /tissue_type="skin"
 /dev_stage="10 days neonate"
 /lab_host="DH10B"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken

derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 107 a 94 c 100 g 97 t
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x BE840995/rev ..
Align seg 1/1 to reverse of: BE840995 from: 1 to: 398

683 GlnGlnThrThrLeuGlyGluValVal 691
|||||
42 CAGCAACGACCTTGGTGAAGTCGTC 16

seq_name: gb_est1:AV625127

seq_documentation_block:

LOCUS AV625127 465 bp mRNA linear EST 15-DEC-2000
DEFINITION AV625127 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone LC088c03_r 5', mRNA sequence.

ACCESSION AV625127
VERSION AV625127.1 GI:10774304
KEYWORDS EST.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Chlamydomonas reinhardtii

REFERENCE Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.

AUTHORS Asamizu, E., Miura, K., Kuch, K., Inoue, Y., Fukuzawa, H., Ohnaka, K., Nakamura, Y., and Tabata, S.

TITLE Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii

JOURNAL DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644

COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES

source

1..465
/organism="Chlamydomonas reinhardtii"
/strain="C9"
/db_xref="taxon:3055"
/clone="LC088c03_r"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from cells cultured in a carbon stress acclimatized condition in which carbon dioxide concentration in the bubbling gas was changed from 5% to 0.04%"

BASE COUNT 118 a 143 c 104 g 100 t
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AV625127 ..

Align seg 1/1 to: AV625127 from: 1 to: 465

49 LeuGlnSerValLeuProPheArgLeu 57
|||||
177 TTGCAATCAGTGTCCCTTCGTTG 203

seq_name: gb_gss:AQ667844

seq_documentation_block:

LOCUS AQ667844 472 bp DNA linear GSS 24-JUN-1999
DEFINITION HS_5383_AL_G02_MR RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-959 Col-3 Row-M, DNA sequence.

ACCESSION AQ667844

VERSION AQ667844.1 GI:5200590

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 472)

Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Kellar, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering.bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu

Plate: 959 row: M column: 3

Seq primer: T7

Class: BAC ends

High quality sequence stop: 472.

FEATURES

source

1..472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-959 Col-3 Row-M"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"

BASE COUNT 123 a 118 c 89 g 141 t
ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AQ667844 ..

Align seg 1/1 to: AQ667844 from: 1 to: 472

585 GlyTrpAsnTyrSerSerLeuAspArg 593

|||||

212 GGATGGAACATATAGCTCTCTGGATAGA 238

seq_name: gb_gss:AQ837806

seq_documentation_block:

LOCUS AQ837806 482 bp DNA linear GSS 30-AUG-1999
DEFINITION HS_5521_A2_F12_SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=1097 Col=24 Row=K, DNA sequence.

ACCESSION AQ837806
VERSION AQ837806.1 GI:5807680
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 482)

REFERENCE Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
AUTHORS

TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT

Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center

University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(piederdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1097 row: K column: 24
Seq primer: SP6
Class: BAC ends

High quality sequence stop: 482.

FEATURES Location/Qualifiers

source

1..482
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=1097 Col=24 Row=K"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

BASE COUNT 185 a 64 c 86 g 139 t 8 others

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AQ837806/rev ..

Align seg 1/1 to reverse of: AQ837806 from: 1 to: 482

786 LeuSerIleSerTyrAlaIysProLeu 794

|||||
172 TTGAGCATCTCATATGCTAAACCAATTA 146

seq_name: gb_gss:A2836677

seq_documentation_block:

LOCUS A2836677 566 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0131D07R Mouse 10kb plasmid UUGC1M library Mus musculus genomic

clone UUGC2M0131D07 R, DNA sequence.

ACCESSION A2836677
VERSION A2836677.1 GI:13006585
KEYWORDS house mouse.
SOURCE Mus musculus.

ORGANISM Mus musculus.

REFERENCE

AUTHORS

1 (bases 1 to 566)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.

Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

Unpublished (2000)

JOURNAL

COMMENT

Contact: Robert B. Weiss

University of Utah Genome Center

University of Utah

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0131 row: D column: 07

Seq primer: CACACAGGAAACAGCTATGACC

Class: plasmid ends

High quality sequence stop: 566.

FEATURES

source

1..566
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0131D07"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptor DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
adaptor vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 161 a 128 c 135 g 142 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x A2836677/rev ..

Align seg 1/1 to reverse of: A2836677 from: 1 to: 566

61 ValSerGluAsnGlnLeuAlaAspGly 69

|||||

180 GTTTCAGAAAACAGCTAGCAGATGGT 154

seq_name: gb_gss:BH025710

seq_documentation_block:

LOCUS BH025710 566 bp DNA linear GSS 17-JUL-2001
 DEFINITION RPCI-24-259D20.TV RPCI-24 Mus musculus genomic clone RPCI-24-259D20

, DNA sequence.

ACCESSION BH025710

VERSION BH025710.1 GI:14789174

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 566)

Zhao,S., Nierman,W., Malek,J., Shvartsman,S., Akinret,B., Levins,M.,

Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,

Russell,D., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-24

Unpublished (1999)

Other_GSSs: RPCI-24-259D20.TJ

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong

(pdejong@mail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.chori.org/bacpac/orderingframe.htm>). BAC end

page: http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html

Plate: 259 row: D column: 20

Seq primer: T7

Class: BAC ends.

FEATURES

source

1..566

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-24-259D20"

/clone_lib="RPCI-24"

/sex="Male"

/cell_type="Spleen/Brain"

/note="Vector: pTARBAC1; Site.1: BamH1; Site.2: BamH1;

RPCI-24 Mouse BAC Library produced by Pieter de Jong. The

library was cloned in the pTARBAC1 cloning vector at the

BamH1 sites using MboI partially digested male C57BL/6J

DNA."

BASE COUNT 125 a 116 c 181 g 144 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x BH025710 ..

Align seg 1/1 to: BH025710 from: 1 to: 566

367 GlnAspGluValLeuArgArgGluMet 375

|||||

177 CAGATGAGGCTTAAAGAGACAGATG 203

seq_name: gb_gss:AQ509377

seq_documentation_block:

LOCUS AQ509377

592 bp DNA linear GSS 04-MAY-1999

DEFINITION

nbxb0096K15r CUGI Rice BAC Library Oryza sativa genomic clone

nbxb0096K15r, DNA sequence.

ACCESSION AQ509377

VERSION AQ509377.1 GI:4733460

KEYWORDS GSS.

SOURCE Oryza sativa.

ORGANISM

Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartioideae; Oryzeae; Oryza.

1 (bases 1 to 592)

Wing,R.A. and Dean,R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: GGAACAGCTATGACCATG

Class: BAC ends

High quality sequence stop: 276.

FEATURES

source

1..592

/organism="Oryza sativa"

/strain="Japonica"

/cultivar="Nipponbare"

/db_xref="taxon:4530"

/clone="nbxb0096K15r"

/clone_lib="CUGI Rice BAC Library"

/tissue_type="Leaf"

/lab_host="E. coli DH10B"

/note="Vector: pBelobAC11; Site.1: HindIII; Site.2:

HindIII; Rice is one of two most popular grains in the

world. Half of the world population especially those

inhabiting highly populated areas of the humid tropics

and subtropics, rely on rice as their primary source of

carbohydrate. Monocotyledonous rice is a diploid plant

(2n=24) with a haploid genome equivalent of 431 Mbp

(Arumuganathan and Earle, 1991). The relatively small

genome of rice, three times larger than that of

Arabidopsis, makes it suitable for genomic studies. In

order to facilitate positional cloning, physical mapping

and genome sequencing of rice, we have constructed a BAC

library from Oryza sativa, Nipponbare variety. The

library contains 36,864 clones with an average insert size

of 128.5 Kb providing 10.9 haploid genome equivalents. The

deep coverage allows the isolation a particular sequence

with a probability of 99.9 %. Two high density filters,

each containing 18,432 clones (doubly spotted), represent

the whole library for colony screening."

BASE COUNT 172 a 111 c 118 g 189 t

ORIGIN

alignment_scores:

Quality: 9.00 Length: 9

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AQ509377/rev ..

Align seg 1/1 to reverse of: AQ509377 from: 1 to: 592

507 IleSerAsnTyrValLeuAspSerTyr 515

|||||

435 ATATCAAAATTATGACTAGATAGCTAT 409

seq_name: gb_gss:AQ290748


```

seq_documentation_block:
LOCUS      AQ290748                611 bp    DNA        linear        GSS 03-DEC-1998
DEFINITION      nbxb0037L03f CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION      AQ290748
VERSION        AQ290748.1  GI:3952110
KEYWORDS       GSS.
SOURCE         Oryza sativa.
ORGANISM       Oryza sativa.
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
               Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE      1 (bases 1 to 611)
AUTHORS       Wing, R.A. and Dean, R.A.
TITLE         A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL       Unpublished (1998)
COMMENT       Contact: Wing RA
               Clemson University Genomics Institute
               Clemson University
               100 Jordan Hall, Clemson, SC 29634, USA
               Tel: 864 656 7288
               Fax: 864 656 4293
               Email: twing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence stop: 376.

```

```

FEATURES             source
     1..611
     /organism="Oryza sativa"
     /strain="Japonica"
     /cultivar="Nipponbare"
     /db_xref="taxon:4530"
     /clone="nbxb0037L03f"
     /clone_lib="CUGI Rice BAC Library"
     /tissue_type="Leaf"
     /lab_host="E. coli DH10B"
     /note="Vector: pBelOBAC11; Site.1: HindIII; Site.2:
HindIII; Rice is one of two most popular grains in the
world. Half of the world population especially those
inhabiting highly populated areas of the humid tropics
and subtropics, rely on rice as their primary source of
carbohydrate. Monocotyledonous rice is a diploid plant
(2n=24) with a haploid genome equivalent of 431 Mbp
(Arumuganathan and Earle, 1991). The relatively small
genome of rice, three times larger than that of
Arabidopsis, makes it suitable for genomic studies. In
order to facilitate positional cloning, physical mapping
and genome sequencing of rice, we have constructed a BAC
library from Oryza sativa, Nipponbare variety. The
library contains 36,864 clones with an average insert size
of 128.5 Kb providing 10.9 haploid genome equivalents. The
deep coverage allows the isolation a particular sequence
with a probability of 99.9 %. Two high density filters,
each containing 18,432 clones (doubly spotted), represent
the whole library for colony screening."
BASE COUNT      183 a   112 c   120 g   196 t
ORIGIN

```

```

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
US-09-701-711-2 x AQ290748/rev ..

```

```

Align seg 1/1 to reverse of: AQ290748 from: 1 to: 611

```

```

. 507 lIeSerAsnTyrValLeuAspSerTyr 515
|||||
435 ATATCAATATGACTAGTAT 409

```

```

seq_name: gb_est2:BI873688
seq_documentation_block:
LOCUS      BI873688                627 bp    mRNA        linear        EST 11-OCT-2001
DEFINITION      963110E06.xl C. reinhardtii CC-1690, Stress condition I, normalized
               , Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION      BI873688
VERSION        BI873688.1  GI:16071692
KEYWORDS       EST.
SOURCE         Chlamydomonas reinhardtii.
ORGANISM       Chlamydomonas reinhardtii
               Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
               Chlamydomonadaceae; Chlamydomonas.
REFERENCE      1 (bases 1 to 627)
AUTHORS       Grossman, A., Chang, C.-W., Davies, J., Harris, E., Hauser, C., Lefebvre,
               P., McDermott, J.P., Shrager, J., Silflow, C. and Stern, D.
TITLE         Analyses of the Chlamydomonas reinhardtii Genome: A Model,
               Unicellular System for Analyzing Gene Function and Regulation in
               Vascular Plants. Project: 963
JOURNAL       Unpublished (2001)
COMMENT       Contact: Charles Hauser
               DCMB Box 91000
               Duke University
               Durham, NC 27708-1000
               Tel: 919 613 8159
               Fax: 919 613 8177
               Email: chauser@duke.edu.

```

```

FEATURES             source
     1..627
     /organism="Chlamydomonas reinhardtii"
     /strain="CC-1690 wild type mt+ 21gr"
     /db_xref="taxon:3055"
     /clone_lib="C. reinhardtii CC-1690, Stress condition I,
normalized, Lambda Zap II"
     /note="Vector: pBluescript II SK-; Site.1: EcoRI; Site.2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr
, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda zap II
(Stratagene) in the EcoRI (5') and XhoRI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT      158 a   196 c   136 g   137 t
ORIGIN

```

```

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
US-09-701-711-2 x BI873688 ..

```

```

Align seg 1/1 to: BI873688 from: 1 to: 627

```

```

49 LeuGlnSerValLeuProPheArgLeu 57
|||||
183 TTGCAATCAGTCTCCCTTCGTTG 209

```

```

seq_name: gb_gss:A2391700

```

```

seq_documentation_block:
LOCUS      A2391700                635 bp    DNA        linear        GSS 03-OCT-2000
DEFINITION      IM0154C06F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
               clone UUGCLM0154C06 F, DNA sequence.
ACCESSION      A2391700
VERSION        A2391700.1  GI:10506743

```

Thu Sep 19 10:04:52 2002

us-09-701-711-2.sepl7oli.rst

```

KEYWORDS
SOURCE      house mouse.
ORGANISM    Mus musculus
REFERENCE   1 (bases 1 to 635)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0154 row: C column: 06
            Seq primer: CGTTGTAAACGACGCCAGT
            Class: plasmid ends
            High quality sequence stop: 635.

FEATURES             source
1. .635
   /organism="Mus musculus"
   /strain="C57BL/6J"
   /db_xref="taxon:10090"
   /clone="UUGCLM0154C06"
   /clone_lib="Mouse 10kb plasmid UUGCLM library"
   /sex="Male"
   /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
   /note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      245 a 117 c 147 g 126 t
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AZ391700/rev ..
Align seg 1/1 to reverse of: AZ391700 from: 1 to: 635

701 SerGluLeuLeuLeuProLeuProPhe 709
|||||
503 TCAGAACTGATATGCCCTCCCTTC 477

seq_name: gb_est2:BI998945

GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 635)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0154 row: C column: 06
Seq primer: CGTTGTAAACGACGCCAGT
Class: plasmid ends
High quality sequence stop: 635.

FEATURES             source
1. .635
   /organism="Mus musculus"
   /strain="C57BL/6J"
   /db_xref="taxon:10090"
   /clone="UUGCLM0154C06"
   /clone_lib="Mouse 10kb plasmid UUGCLM library"
   /sex="Male"
   /lab_host="E. Coli strain XL10-Gold, TI-resistant, F-"
   /note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
BASE COUNT      245 a 117 c 147 g 126 t
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AZ391700/rev ..
Align seg 1/1 to reverse of: AZ391700 from: 1 to: 635

701 SerGluLeuLeuLeuProLeuProPhe 709
|||||
503 TCAGAACTGATATGCCCTCCCTTC 477

seq_name: gb_est2:BI998945

seq_documentation_block:
LOCUS      BI998945          649 bp      mRNA      linear      EST 25-OCT-2001
DEFINITION Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BI998945
VERSION   BI998945.1 GI:16433719
KEYWORDS  EST.
SOURCE    Chlamydomonas reinhardtii.
ORGANISM  Chlamydomonas reinhardtii.
REFERENCE 1 (bases 1 to 649)
AUTHORS   Grossman,A., Chang,C.-W., Davies,J., Harris,E., Hauser,C., Lefebvre
,P., McDermott,J.P., Shrager,J., Silflow,C. and Stern,D.
TITLE     Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants. Project: 1031
JOURNAL   Unpublished (2001)
COMMENT   Contact: Charles Hauser
            DCMB Box 91000
            Duke University
            Durham, NC 27708-1000
            Tel: 919 613 8159
            Fax: 919 613 8177
            Email: chauser@duke.edu.
            Location/Qualifiers
            1. .649
              /organism="Chlamydomonas reinhardtii"
              /strain="CC-1690 wild type mt+ 21gr"
              /db_xref="taxon:3055"
              /clone_lib="C. reinhardtii CC-1690, Stress II (normalized
              ), Lambda Zap II"
              /note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2:
              XhoI; Stress condition II library, constructed by John
              Davies and Jeffrey McDermott, combines cDNAs from CC-1690
              cells grown to mid-log phase in TAP (NH4+ - containing)
              and shifted to TAP - NO3- (24hrs); H2 production
              conditions (0, 12hr, 24hr) see Melis et al., (2000) Plant
              Phys. 122: 127-135; TAP + H2O2 (1, 12, 24 hr); TAP +
              sorbitol (1, 2, 6, 24 hr); TAP + Cd (1, 2, 6, 24 hr).
              POLYA mRNA was purified from each sample, pooled and cDNA
              synthesized. The cDNA was directionally cloned into lambda
              Zap II (Stratagene) in the EcoRI (5') and XhoI (3')
              sites. pBluescript II SK- plasmids were excised from the
              lambda Zap clones by superinfection with EXAssist
              (Stratagene) phage. The library was normalized using
              method 4 described in Bonaldo et al., (1996) Genome
              Research 6: 791-806."
              Research 6: 791-806."
BASE COUNT      162 a 207 c 145 g 134 t 1 others
ORIGIN

alignment_scores:
  Quality:      9.00      Length:      9
  Ratio:        1.000     Gaps:      0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BI998945 ..
Align seg 1/1 to: BI998945 from: 1 to: 649

49 LeuGlnSerValLeuProPheArgLeu 57
|||||
183 TTGCAATCAGTGTCCCTTCGTTG 209

seq_name: gb_est2:BG177050

seq_documentation_block:
LOCUS      BG177050          655 bp      mRNA      linear      EST 06-FEB-2001
DEFINITION 602313868F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4419520 5',
mRNA sequence.

```

```

ACCESSION      BG177050
VERSION        BG177050.1  GI:12683753
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 655)
AUTHORS        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE          NTH-MGC http://mgc.nci.nih.gov/.
JOURNAL        National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT        Unpublished (1999)
                Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: Louis Staudt, M.D., Ph.D.
                CDNA Library Preparation: Life Technologies, Inc.
                CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: L1AM10155 row: h column: 17
                High quality sequence stop: 471.
FEATURES       source
                Location/Qualifiers
                1..655
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4419520"
                /clone_lib="NTH-MGC-85"
                /tissue_type="lymphoma, cell line"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: lymph; Vector: pCMV-Sport6; Site 1: NotI;
                Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 1.867 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT    184 a 141 c 112 g 218 t
ORIGIN

alignment_scores
  Quality: 9.00 Length: 9
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block
US-09-701-711-2 x BG177050 ..
Align seg 1/1 to: BG177050 from: 1 to: 655
387 LyslleGlnLeuSerArgAlaArgLeu 395
|||||
256 AGATTTCAGCTCTCTCGTGCAGCTTTG 282

seq_name: gb_gss:CNS04ADV

seq_documentation_block:
LOCUS      CNS04ADV 919 bp DNA linear GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
            095G04 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION  AL281740
VERSION    AL281740.1 GI:8020066
KEYWORDS   GSS: genome survey sequence.
SOURCE     Tetraodon nigroviridis.
ORGANISM   Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE  1 (bases 1 to 919)
AUTHORS    Roest-Crolius.H., Jaillon.O., Dasilva.C., Fizesma.C., Fisher.C.,
            Bouneau.L., Billault.A., Quetier.F., Saurin.W., Bernot.A. and
            Weissenbach.J.

```

```

TITLE          Characterization and repeat analysis of the compact genome of the
                freshwater pufferfish Tetraodon nigroviridis
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 919)
AUTHORS        Roest-Crolius.H., Jaillon.O., Dasilva.C., Bouneau.L., Fisher.C.,
                Bernot.A., Fizesma.C., Wincker.P., Brottier.P., Quetier.F.,
                Saurin.W. and Weissenbach.J.
                Human gene number estimate provided by genome wide analysis using
                Tetraodon nigroviridis DNA sequence
JOURNAL        Unpublished
REFERENCE      3 (bases 1 to 919)
AUTHORS        Genoscope.
TITLE          Direct Submission
JOURNAL        Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT        This sequence is a single read and was generated as part of a large
                scale clone-end sequencing project of the tetraodon nigroviridis
                genome. For more information, please take a look at
                http://www.genoscope.cns.fr/tetraodon.
FEATURES       source
                Location/Qualifiers
                1..919
                /organism="Tetraodon nigroviridis"
                /db_xref="taxon:99883"
                /clone="095G04"
                /clone_lib="G"
                /note="Genoscope sequence ID : C0BG095BD02LPL-end : T7"
BASE COUNT    232 a 212 c 241 g 229 t
ORIGIN

alignment_scores
  Quality: 9.00 Length: 9
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block
US-09-701-711-2 x CNS04ADV ..
Align seg 1/1 to: CNS04ADV from: 1 to: 919
121 LeuLysAsnAlaGlyLeuAlaValGly 129
|||||
536 CTAAAGACGCTGGGTAGCTGTCGGG 562

seq_name: gb_est2:BG397547

seq_documentation_block:
LOCUS      BG397547 1374 bp mRNA linear EST 12-MAR-2001
DEFINITION 602439379f1 NIH_MGC_48 Homo sapiens CDNA clone IMAGE:456586 5',
            mRNA sequence.
ACCESSION  BG397547
VERSION    BG397547.1 GI:13290995
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1374)
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            CDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1CMI282 row: n column: 19
            High quality sequence start: 33
            High quality sequence stop: 261.
FEATURES     Location/Qualifiers

```

```

source
1. .1374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:456586"
/tissue_type="NIH_MGC_48"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library." 1 others
BASE COUNT      299 a 347 c 425 g 302 t
ORIGIN

alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BG420056 ..
Align seg 1/1 to: BG420056 from: 1 to: 1597

437 GlyTyrSerGlnSerGlyGlyValThr 445
|||||
1348 GGTACTCACAAATCGGTGGTGTGACA 1374

seq_name: gb_est1:AW796076

seq_documentation_block:
LOCUS      AW796076                116 bp      mRNA      linear      EST 16-MAY-2000
DEFINITION MR2-UM0025-300300-102-a09 UM0025 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW796076
VERSION    AW796076.1 GI:7847946
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 116)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
            Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-MR2-UM0025-300
300-102-a09&t3=2000-03-30&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 115.
Location/Qualifiers
            1..116
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="UM0025"
            /dev_stage="Adult"
            /note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2:
            SmaI; A mini-library was made by cloning products derived
            from ORESTES PCR (U.S. Letters Patent application No. 196
            716 - Ludwig Institute for Cancer Research) profiles
            into the pUC 18 vector. Reverse transcription of tissue
            mRNA and cDNA amplification were performed under low
            stringency conditions."

FEATURES
source
1. .1597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4591882"
/tissue_type="NIH_MGC_14"
/lab_host="DH10B (phage-resistant)"
/note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',

```

```

1. .1374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:456586"
/tissue_type="NIH_MGC_48"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
Note: this is a NIH_MGC Library." 1 others
BASE COUNT      299 a 347 c 425 g 302 t
ORIGIN

alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BG397547/rev ..
Align seg 1/1 to reverse of: BG397547 from: 1 to: 1374

461 HisValAsnAlaSerPheSerArgSer 469
|||||
175 CATGTCATCAAGCTTTTCCCGATCC 149

seq_name: gb_est2:BG420056

seq_documentation_block:
LOCUS      BG420056                1597 bp      mRNA      linear      EST 14-MAR-2001
DEFINITION 602453743F1 NIH_MGC_14 Homo sapiens cDNA clone IMAGE:4591882 5',
            mRNA sequence.
ACCESSION  BG420056
VERSION    BG420056.1 GI:13326562
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1597)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: Incyte Genomics, Inc.
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
Plate: LLCMI329 row: f column: 11
High quality sequence stop: 3.
Location/Qualifiers
            1..1597
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4591882"
            /clone_lib="NIH_MGC_14"
            /tissue_type="renal cell adenocarcinoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: kidney; Vector: pOTB7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5',

```

BASE COUNT 22 a 41 c 36 g 17 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AW796076

Align seg 1/1 to: AW796076 from: 1 to: 116

662 ArgGlyTyrAspGlnSerSerLeu 669

|||||
34 CGCGGGTATGACCATCCAGCCTA 57

seq_name: gb_gss:AZ286261

seq_documentation_block:

LOCUS AZ286261 119 bp DNA linear GSS 27-JUL-2000
DEFINITION RPCI-23-442122.TJ RPCI-23 Mus musculus genomic clone RPCI-23-442122
, DNA sequence.

ACCESSION AZ286261
VERSION A2286261.1 GI:9528137

KEYWORDS GSS.
SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 119)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret

,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other_GSSs: RPCI-23-442122.TV
Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics ([info@resgen.com](http://www.resgen.com)). BAC end page:

http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 442 row: I column: 22
Seq primer: SP6

Class: BAC ends.

FEATURES

source

1. .119
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-442122"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 18 a 30 c 39 g 32 t
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AZ286261/rev

Align seg 1/1 to reverse of: AZ286261 from: 1 to: 119

177 AlaGlyLysProAlaArgVal 184

|||||
77 GCAGAGGCGACCGCCAGGCTA 54

seq_name: gb_gss:AZ909005

seq_documentation_block:

LOCUS AZ909005 136 bp DNA linear GSS 05-MAR-2001
DEFINITION RPCI-24-220018.TJ RPCI-24 Mus musculus genomic clone RPCI-24-220018
, DNA sequence.

ACCESSION AZ909005
VERSION A2909005.1 GI:132727950

KEYWORDS GSS.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 136)
AUTHORS Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-24
Unpublished (1999)

Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200

Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-24. For BAC
library availability, please contact Pieter de Jong

(pdejong@avail.cho.org). Clones may be purchased from BACPAC

Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end

page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 220 row: O column: 18
Seq primer: SP6

Class: BAC ends.

FEATURES

source

1. .136
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-220018"
/clone_lib="RPCI-24"
/sex="Male"
/cell_type="Spleen/Brain"

/note="Vector: pPARBAC1; Site_1: BamHI; Site_2: BamHI;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pPARBAC1 cloning vector at the
BamHI sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 34 a 32 c 46 g 24 t
ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

```

US-09-701-711-2 x AZ909005
Align seg 1/1 to: AZ909005 from: 1 to: 136

769 LeuArgTyrSerAlaGlyValGly 776
|||||
50 CTCAGATACACTACGCCGGTGAGGG 73

seq_name: gb_est2:BF983453

seq_documentation_block:
LOCUS BF983453 203 bp mRNA linear EST 23-JAN-2001
DEFINITION 602307077F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4398427 5',
mRNA sequence.
ACCESSION BF983453
VERSION BF983453.1 GI:12386265
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 203)
NIH-MGC http://mhc.nci.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10100 row: i column: 20
High quality sequence stop: 203.
Location/Qualifiers
1..203
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4398427"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 20 a 65 c 94 g 24 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BF983453
Align seg 1/1 to: BF983453 from: 1 to: 203

667 SerSerLeuGlyProArgSerGln 674
|||||
95 AGCTCGCTCGTCCAGGTCCTCCAG 118

seq_name: gb_est1:AV257342

seq_documentation_block:
LOCUS AV257342 206 bp mRNA linear EST 04-NOV-1999
DEFINITION AV257342 RIKEN full-length enriched, adult male testis (DH10B) Mus
musculus cDNA clone 4921535N13 3' similar to AF053232 Mus musculus

```

```

SIK similar protein mRNA, mRNA sequence.
AV257342
AV257342.1 GI:6244801
EST.
house mouse.
Mus musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 206)
Konno.H., Aizawa.K., Akahira.S., Akiyama.J., Carninci.P., Endo.T.,
Fukuda.S., Fukunishi.Y., Hara.A., Hayatsu.N., Hirozane.T., Hori.F.,
Ishii.Y., Ishikawa.T., Itoh.M., Izawa.M., Kadota.K., Kagawa.I., Kai
C., Kawai.J., Kikuchi.N., Kojima.Y., Koya.S., Kusakabe.M., Kai
Matsuyama.T., Miki.R., Mizuno.Y., Nakamura.M., Oda.H., Okazaki.Y.,
Owa.C., Ozawa.Y., Saito.H., Sano.M., Sato.K., Shibata.K., Shibata
Y., Shigemoto.Y., Shiraki.T., Sogabe.Y., Sugahara.Y., Suzuki.H.,
Suzuki.H., Takahashi.F., Tateo.M., Tominaga.N., Tsunoda.Y.,
Watahiki.A., Watanabe.S., Yamamura.T., Yasunishi.A., Yokota.T.,
Yoshiki.A., Yoshino.M., Muramatsu.M. and Hayashizaki.Y.
RIKEN Mouse ESTs (Konno.H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Sasaki.N., Izawa.M., Watahiki.M., Ozawa.K., Tanaka.T., Yoneda.Y.,
Matsuura.S., Carninci.P., Muramatsu.M., Okazaki.Y. and Hayashizaki
Y.
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh.M., Kitsunai.T., Akiyama.J., Shibata.K., Izawa.M., Kawai.J.,
Tomaru.Y., Carninci.P., Shibata.Y., Ozawa.Y., Muramatsu.M., Okazaki
Y. and Hayashizaki.Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci.P. and Hayashizaki.Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
Location/Qualifiers
1..206
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4921535N13"
/clone_lib="RIKEN full-length enriched, adult male testis
(DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGAGGATCCAGAGCTCTTTTTTTTTTTTTTTT 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCGAGTTAATTAATTAATTAATCCCCCCCCCCCCC 3']. cDNA
was cloned into the xhoI and BamHI sites. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC 1. Cloning sites, 5' end: SalI; 3' end: BamHI."

```

BASE COUNT 52 a 53 c 32 g 69 t

plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

BASE COUNT 38 a 39 c 91 g 38 t 2 others
ORIGIN

alignment_block:
US-09-701-711-2 x AV257342/rev ..

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

Align seg 1/1 to reverse of: AV257342 from: 1 to: 206

seq_name: gb_est2:BM128835

Align seg 1/1 to reverse of: BM128835 from: 1 to: 208

227 ThrSerLeuGluAsnLeuArgAla 234
|||||
116 ACCTCCCTTGAAATTTGAGAGCA 93

591 LeuAspArgProValPheProThr 598
|||||
88 CTGGACCGACCGTCTTCCCCACG 65

seq_name: gb_est2:R70943

seq_documentation_block:
LOCUS R70943 215 bp mRNA linear EST 01-JUN-1995
DEFINITION y150e05.s1 Soares placenta Nb2HP Homo sapiens cDNA clone
IMAGE:142688 3', mRNA sequence.

ACCESSION BM128835
VERSION BM128835.1 GI:17123387
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 208)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K., Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T., Jackson, Y., and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brownjfas.harvard.edu)
Trace considered overall poor quality
High quality sequence stop: 1.

REFERENCE 1 (bases 1 to 215)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.

TITLE
JOURNAL
COMMENT
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1177
High quality sequence stops: 161
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1177 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 161.

Location/Qualifiers
1. .215
/organism="Homo sapiens"
/db_xref="GDB:551767"
/db_xref="taxon:9606"
/clone="IMAGE:142688"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="Dh10B"
/lab_host="placenta obtained at birth (full term)"
/notes="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAAATTCGGCGCCGAGGAAATTTTTTTTTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

FEATURES
source
1. .215
/organism="Homo sapiens"
/db_xref="GDB:551767"
/db_xref="taxon:9606"
/clone="IMAGE:142688"
/clone_lib="Soares placenta Nb2HP"
/sex="Female"
/dev_stage="placenta obtained at birth (full term)"
/lab_host="Dh10B (ampicillin resistant)"
/notes="Organ: placenta; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGAAATTCGGCGCCGAGGAAATTTTTTTTTTTT 3', double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

FEATURES
source
1. .208
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="Dh10B"

/notes="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal I; Starting library constructed using SuperScript plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library

and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

BASE COUNT
ORIGIN

79 a 37 c 24 g 72 t 3 others

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x R70943 ..

Align seg 1/1 to: R70943 from: 1 to: 215

703 Leu1leuProLeuPropheLys 710

|||||

147 CTTATACCTCCACTTCCTTTAAA 170

seq_name: gb_est2:BF377641

seq_documentation_block:
LOCUS BF377641 217 bp mRNA linear EST 24-NOV-2000
DEFINITION PM1-TN0120-270800-001-al2 TN0120 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF377641

VERSION BF377641.1 GI:11339666

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 217)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=PM1st2=PM1-TN0120-270800-001-al2&t3=2000-08-27&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 2

High quality sequence stop: 217.

Location/Qualifiers

FEATURES

source

1..217

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="TN0120"

/dev_stage="Adult"

/note="Organ: testis.normal; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 74 a 32 c 40 g 71 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x BF377641 ..

Align seg 1/1 to: BF377641 from: 1 to: 217

250 LysLeuAsnIleAsnGluAspLys 257

|||||

162 AAGTTGAATATTATGAGGATAAA 185

seq_name: gb_est2:BI493603

seq_documentation_block:

LOCUS BI493603 233 bp mRNA linear EST 28-AUG-2001

DEFINITION df103c08.w1 Morton Fetal Cochlea Homo sapiens cDNA clone

IMAGE:2540942 3', mRNA sequence.

ACCESSION BI493603

VERSION BI493603.1 GI:15332947

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 233)

AUTHORS Robertson,N.G., Khertarpal,U., Gutierrez-Espeleta,G.A., Bieber,F.R. and Morton,C.C.

TITLE Isolation of novel and known genes from a human fetal cochlear cDNA

library using subtractive hybridization and differential screening

Genomics 23, 42-50 (1994)

JOURNAL 95130111

MEDLINE

COMMENT

Contact: Morton, C. C.

Departments of Pathology and Obstetrics, Gynecology and

Reproductive Biology

Brigham and Women's Hospital

75 Francis Street, Harvard Medical School, Boston, MA 02115, USA

Tel: 617 732 7980

Fax: 617 738 6996

Email: cmorton@bics.bwh.harvard.edu

DNA sequencing and analyses were performed by National Institutes

of Health Intramural Sequencing Center (NISC; see

http://www.nisc.nih.gov).

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Plate: LLAM6329 row: F column: 15

Seq primer: T7 primer.

Location/Qualifiers

1..233

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:2540942"

/tissue_type="Cochlea"

/dev_stage="16-22 week fetus"

/lab_host="SOLR cells (kanamycin resistant)"

/note="Organ: ear; Vector: pBluescript SK-; Site_1: EcoRI;

Site_2: XhoI; Reference: Genomics 23, 42-50 (1994) Cloned.

37% of inserts <0.5 kb, 56% 0.5-1.0 kb, 7% >1 kb. Uni-ZAP

XR Vector. Library constructed by N. Robertson, C. Morton.

-5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor

sequence: 5' CTCGATTTTTTTTTTTTTTTT 3'"

BASE COUNT 38 a 70 c 65 g 60 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x BI493603/rev ..

Align seg 1/1 to reverse of: BI493603 from: 1 to: 233

113 ProLysGluGlyLeuGlnGluGly 120
 218 CCCAAGAAGGCTCCAGAAGGC 195

seq_name: gb_estl:AA729757

seq_documentation_block:
 LOCUS AA729757 244 bp mRNA linear EST 27-JAN-1998
 DEFINITION nx61d08.s1 NCI_CGAP_Al1v1 Homo sapiens cDNA clone IMAGE:1266735,
 mRNA sequence.
 ACCESSION AA729757
 VERSION AA729757.1 GI:2751116
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 244)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 CDNA Library Preparation: David B. Krizman, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone Distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 349 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

source
 1..244
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1266735"
 /clone_lib="NCI_CGAP_Al1v1"
 /tissue_type="alveolar rhabdomyosarcoma"
 /lab_host="DH10B"
 /note="Vector: pAMP10; mRNA made from alveolar
 rhabdomyosarcoma, cDNA made by oligo-dT priming.
 Non-directionally cloned. Size-selected on agarose gel,
 average insert size 600 bp. Reference: Krizman et al.
 (1996) Cancer Research 56:5380-5383."
 BASE COUNT 93 a 27 c 42 g 82 t
 ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AA729757/rev ..

Align seg 1/1 to reverse of: AA729757 from: 1 to: 244

703 LeuLeuProLeuProPheLys 710
 36 CTTATACCTCCACTCCCTTTAAA 13

seq_name: gb_estl:BB417742

seq_documentation_block:

LOCUS BB417742 253 bp mRNA linear EST 16-JUL-2000
 DEFINITION BB417742 RIKEN full-length enriched, 12 days embryo spinal cord Mus
 musculus cDNA clone C530005A18 3', mRNA sequence.
 ACCESSION BB417742.1 GI:9239097
 VERSION BB417742
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 253)
 AUTHORS Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
 P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
 Hirozane,T., Hori,Y., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
 Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
 Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,
 Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugihara,Y.,
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toya
 T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
 Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
 M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Konno,H., et al.)

TITLE

JOURNAL

COMMENT

Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-resesc.riken.go.jp,

URL:http://genome.gsc.riken.go.jp/

Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Thermotabilization and thermoactivation of thermostable enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh,M., Kitsumai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
 Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)

Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.

Location/Qualifiers
 1..253
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="C530005A18"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 spinal cord"
 /tissue_type="spinal cord"
 /dev_stage="12 days embryo"
 /lab_host="DH10B"

FEATURES

source

1..253
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="C530005A18"
 /clone_lib="RIKEN full-length enriched, 12 days embryo
 spinal cord"
 /tissue_type="spinal cord"
 /dev_stage="12 days embryo"
 /lab_host="DH10B"
 /note="Site_1: Sali; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'-
 GAGAGAGAAGGATCCAGAGCTCTTTTTTTTTTTTTTTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by

cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTAATAATTCCTCCGCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 69 a 68 c 55 g 61 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BB417742/rev ..

Align seg 1/1 to reverse of: BB417742 from: 1 to: 253

283 LeuThrTyrThrGlnAlaGluLeu 290
|||||
210 CTAACCTACACACAGGCTGAATC 187

seq_name: gb_est1:BB579201

seq_documentation_block:
LOCUS BB579201 258 bp mRNA linear EST 30-NOV-2000
DEFINITION BB579201 RIKEN full-length enriched, 12 days embryo female mullerian duct Mus musculus cDNA clone 6820401C24 5', mRNA sequence.

ACCESSION BB579201
VERSION BB579201.1 GI:11475148
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 258)

Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P., Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodojima,Y., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Konno H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C., Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka ,T., Toya,T., Watahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
Unpublished (2000)

TITLE Yoshihide Hayashizaki
JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
COMMENT The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)

Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki ,Y. and Hayashizaki,Y.

Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)

Carninci,P. and Hayashizaki,Y.

High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)

Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
source

Location/Qualifiers
1. .258
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6820401C24"
/lab_host="RIKEN full-length enriched, 12 days embryo female mullerian duct"
/sex="female"
/tissue_type="mullerian duct includes surrounding region"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/note="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5', GAGAGAGAGATCTCGAGTTAATAATTCCTCCGCCCCCCCCC 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGATCTCGAGTTAATAATTCCTCCGCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI"

BASE COUNT 93 a 42 c 77 g 46 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BB579201 ..

Align seg 1/1 to: BB579201 from: 1 to: 258

230 GluAsnLeuArgAlaLysTyrLeu 237
|||||
155 GAAACCTTACGGCAAGTACTTA 178

seq_name: gb_est1:AA247205

seq_documentation_block:
LOCUS AA247205 262 bp mRNA linear EST 11-MAR-1997
DEFINITION csg0312.seq.F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA 5', mRNA sequence.

ACCESSION AA247205
VERSION AA247205.1 GI:18785990
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 262)

REFERENCE
AUTHORS Liew,C.C.
TITLE cDNAs from human fetal heart (1997)
JOURNAL Unpublished (1997)
COMMENT Contact: Liew CC
Brigham and Women's Hospital
Harvard Medical School
75 Francis St. Boston, MA 02115, USA
Tel: 6177328915
Fax: 6179750995
Email: cliew@rics.bwh.harvard.edu
PCR Primers
FORWARD: 5' GCCAAGCTCGAATTAACCCCTCACTAAGGG 3'
BACKWARD: 5' CCAGTGAATTCATACGACTCACTATAGGG 3'

Seq primer: 5' GAAATTAACCTCTACTAAAGG 3'.

```

FEATURES
  source
    1..262
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="Human fetal heart, Lambda ZAP Express"
      /lab_host="E. coli XL1-Blue"
      /note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
XhoI; mRNA was purified from human fetal hearts (8-10
weeks). cDNA was synthesized using a XhoI-Oligo dT
adaptor-primer. EcoRI adaptors were ligated, followed by
digestion with XhoI, for directional cloning into
predigested lambda ZAP Express."
    70 a 58 c 63 g 71 t
      BASE COUNT
      ORIGIN

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
  US-09-701-711-2 x AA247205 ..
  Align seg 1/1 to: AA247205 from: 1 to: 262

693 GlyAsnAlaLeuAlaThrPheGly 700
233 GCCAATGCATTGGCAACTTTGGT 256

seq_name: gb_est1:AI866213

seq_documentation_block:
  LOCUS AI866213 268 bp mRNA linear EST 07-MAR-2000
  DEFINITION W127c10.x1 NCI_CGAP_Utl1 Homo sapiens cDNA clone IMAGE:2426130 3',
  mRNA sequence.
  ACCESSION AI866213.1 GI:5530320
  VERSION AI866213.1
  KEYWORDS EST.
  SOURCE human.
  ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  REFERENCE 1 (bases 1 to 268)
  AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  JOURNAL Unpublished (1997)
  COMMENT Contact: Robert Strausberg, Ph.D.
  Email: cgaps-r@mail.nih.gov
  Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
  Emmert-Buck, M.D., Ph.D.
  CDNA Library Preparation: Life Technologies, Inc.
  CDNA Library Arrayed by: Greg Lennon, Ph.D.
  DNA Sequencing by: Washington University Genome Sequencing Center
  Clone distribution: NCI-CGAP clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  www-bio.llnl.gov/bbrp/image/image.html
  Insert length: 855 Std Error: 0.00
  Seq primer: -400P from Gibco
  High quality sequence stop: 211.
  Location/Qualifiers
    1..268
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone_lib="NCI_CGAP_Utl1"
      /tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
      /lab_host="DH10B"
      /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

```

Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"

BASE COUNT 98 a 46 c 27 g 97 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AI866213 ..

Align seg 1/1 to: AI866213 from: 1 to: 268

703 LeuIleLeuProLeuProPhelys 710
|||||
198 CTTATACTTCACATCCCTTAA 221

seq_name: gb_est1:AI583432

seq_documentation_block:

LOCUS AI583432 274 bp mRNA linear EST 14-DEC-1999
DEFINITION tr-96h06.x1 NCI_CGAP_Panl Homo sapiens cDNA clone IMAGE:2226971 3',
mRNA sequence.

ACCESSION AI583432
VERSION AI583432.1 GI:4569329
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 274)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert length: 2636 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 197

POLYA-No.

Location/Qualifiers

1..274

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="NCI_CGAP_Panl"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.72 kb. Life Technologies catalog #:

11548-013"

BASE COUNT 97 a 43 c 27 g 107 t

ORIGIN

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-701-711-2 x AI583432 ..

Align seg 1/1 to: A1583432 from: 1 to: 274

703 Leu1leLeuProLeuPropheLys 710 linear EST 31-JUL-2000
|||||
208 CTTATACCTCCACTTCCTCCCTTAA 231

seq_name: gb_estl:BB537706

seq_documentation_block:
LOCUS BB537706 277 bp mRNA linear EST 31-JUL-2000
DEFINITION BB537706 RIKEN full-length enriched, 0 day neonate eyeball Mus
musculus cDNA clone El30007F02 3', mRNA sequence.
ACCESSION BB537706
VERSION BB537706.1 GI:9593206
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 277)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T., Carninci
P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Itoh,M.,
Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.
Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata
Y., Shigemoto,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Sugahara,Y.,
Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomingaga,N., Toya
T., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yamanaka,I.,
Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al.)
Unpublished (2000)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: <http://genome-gsc.riken.go.jp/>
Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S., Sasaki
N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Thermostabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh,M., Kitsuai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for
further details.
Location/Qualifiers
1..277
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="E130007F02"
/clone.lib="RIKEN full-length enriched, 0 day neonate
eyeball"
/tissue_type="eyeball"
/dev_stage="0 day neonate"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCAACTCGAGTTTTTTTTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGATTCTCGAGTTAAATTAATCCCTCCCTCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pluascript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 78 a 77 c 52 g 70 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x BB537706/rev ..
Align seg 1/1 to reverse of: BB537706 from: 1 to: 277

54 PropheArgLeuGlyGlnValval 61
|||||
53 CCATTAGGCTTGCGCAAGTGGTT 30

seq_name: gb_gss:A2375019

seq_documentation_block:
LOCUS A2375019 277 bp DNA linear GSS 02-OCT-2000
DEFINITION IM0128L10F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0128L10 F, DNA sequence.
ACCESSION A2375019
VERSION A2375019.1 GI:10488719
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 277)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0128 row: L column: 10
Seq primer: CGTTGTAACGACGCGCCAGT
Class: plasmid ends
High quality sequence stop: 277.
Location/Qualifiers
1..277
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0128L10"
/clone.lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Tl-resistant, P-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource

(<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114(gb)AF129072.1), a copy-number inducible derivative of plasmid RL. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 67 a 60 c 55 g 95 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AZ375019 ..
Align seg 1/1 to: AZ375019 from: 1 to: 277

703 LeuileLeuProLeuProPhelys 710
|||||
124 CTAATCTTGCCTTTCCTTCAG 147

seq_name: gb_est2:H82436

seq_documentation_block:
LOCUS H82436 293 bp mRNA linear EST 09-NOV-1995
DEFINITION YU80604.s1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone IMAGE:240150 3', mRNA sequence.

ACCESSION H82436
VERSION H82436.1 GI:1060525
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 293)
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project

TITLE Unpublished (1995)
JOURNAL Contact: Wilson RK
COMMENT Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

Insert Size: 1189
High quality sequence stops: 281
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1189 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 281.
Location/Qualifiers

FEATURES
source 1..293
/organism="Homo sapiens"
/db_xref="GDB:3789103"

/db_xref="taxon:9606"
/clone="IMAGE:240150"
/clone_lib="Soares fetal liver spleen lNFLS"
/sex="male"

/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site: 1: Pac I; Site 2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAGAAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 93 a 51 c 41 g 102 t
ORIGIN

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x H82436 ..
Align seg 1/1 to: H82436 from: 1 to: 293

703 LeuileLeuProLeuProPhelys 710
|||||
186 CTTATACTCCACTTCCTTTAAG 209

seq_name: gb_est1:AV205177

seq_documentation_block:
LOCUS AV205177 295 bp mRNA linear EST 30-OCT-1999
DEFINITION AV205177 RIKEN full-length enriched, adult male testis Mus musculus cDNA clone 1700073108 3', mRNA sequence.

ACCESSION AV205177
VERSION AV205177.1 GI:6146030
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 295)
AUTHORS Konno, H., Alzawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tatenno, M., Tominaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Konno, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp
URL: <http://genome.gsc.riken.go.jp/>
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Konno, H., et al. 1999)
JOURNAL Unpublished (1999)
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp
URL: <http://genome.gsc.riken.go.jp/>
Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)

```

BASE COUNT      82 a   63 c   74 g   76 t
ORIGIN

alignment_scores:
    Quality:      8.00          Length:      8
    Ratio:        1.000         Gaps:      0
    Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
US-09-701-711-2 x AV205177/rev ..

Align seg 1/1 to reverse of: AV205177 from: 1 to: 295

772 SerAlaGlyValGlyAlaThrTrp 779
|||||
24 TCTGCTGGGGTGGTGCCACTTGG 1

seq_name: gb_est1:AU100452

seq_documentation_block:
LOCUS      AU100452              300 bp      mRNA      linear      EST 05-APR-2001
DEFINITION AU100452 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
            kala0773 similar to Homo sapiens mRNA for histone H2B clone
            pJG4-5-15, mRNA sequence.
ACCESSION  AU100452
VERSION    AU100452.1 GI:13551581
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  1 (bases 1 to 300)
AUTHORS   Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J., Hatai,
            ,H., Ota,T., Isogai,T., Tanaka,T., Nakamura,Y., Morishita,S., Okubo,
            ,K., Suyama,A. and Sugano,S.
TITLE     In silico mapping of the 5'-ends of human mRNAs using full-length
            enriched and 5'-end enriched cDNA libraries constructed by

```

```

FEATURES
  source
    1. 302
      /organism="Mus musculus"
      /strain="C57BL/6"
      /db_xref="taxon:10090"
      /clone="IMAGE:607026"
      /clone_lib="Stratagene mouse skin (#937313)"
      /sex="females"
      /tissue_type="whole skin"

```

```
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo
dr. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
```

```
BASE COUNT      64 a      70 c      60 g      108 t
ORIGIN
```

```
alignment_scores:
  Quality:      8.00      Length:      8
  Ratio:      1.000      Gaps:      0
Percent Similarity: 100.000      Percent Identity: 100.000
```

```
alignment_block:
US-09-701-711-2 x A1528822/rev ..
Align seg 1/1 to reverse of: A1528822 from: 1 to: 302
```

```
705 LeuProLeuProPheLysGlyAsp 712
|||||
90 CTGCCCCCTACCCCTTCAGGAGAT 67
```


1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17
|||||
1 ATCGCTAATTCATATTTAAAGGTTTTCAGGTCAGTCAATGACATGGC 50
17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaa 34
|||||
51 TGTCTATGATGTAATGTCAACTCATGCAACAAGCGCGGATTTTATGGCAA 100
34 snAspIleThrIleThrGlnArgValThrIleGluSerLeuGln 50
|||||
101 ATGACATTTACCATCACAGACTACAGCGAGTGACCATTTGAAGCTTACAA 150
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67
|||||
151 AGCGTGTCGCGGTTTCGCTTGGGTCAGGTGAGCGAAACCAAGTTGGC 200
67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84
|||||
201 TGATGGTGTCAAGACACTTTATGCAACAGGCAATTTTTCAGATGTGCAAG 250
84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100
|||||
251 TCTATCATCAAGAGGCGTATCATCTATCAGGTAACCGAAGCGCGTTA 300
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117
|||||
301 ATCGCTGAGATTAATTTTGGGGCAATCGCTTAATTTCCAAAGAAGGTCT 350
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134
|||||
351 ACAAGAAGGGCTAAATAATGCTGGCTTACGTGGGTCAACCCTAAAC 400
134 InAlaThrValGlnMetIleGlnThrGluLeuThrAsnGlnTyrIleSer 150
|||||
401 AAGCCACAGTACAGATGATCAACACCGAGCTTACCAATCAATATATATCA 450
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGl 167
|||||
451 CAAGGCTATTATAACCGAAATTAAGCTTGTGCTGAAGGTAACCTGCACGG 500
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184
|||||
501 TAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAACCTGCACGG 550
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
|||||
551 TGGTTGATATTAATATCATTCGCAATCAGCATTTTAGCGATGCAGATTG 600
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217
|||||
601 ATTGATGTGCTTGGCATTAAGGATAATAAATCAATCCACTGTCTAAAGC 650
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234
|||||
651 TGACGGTTATACTCAAGAAAAGCTGGTGACCATTTAGAGAAATTTGGCGTG 700
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250
|||||
701 CTAATATCTCAATGCAGGGTTTGGCGTTTGGATTAAGATGCTTAAG 750
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267
|||||
751 CTTAATATTAAATGAAGATAAAACCGTATCTTTGTTGAGATTTTCATTGCA 800
267 sGluGlyGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
|||||
801 TGAAGGTGAGCAATATCGCTTGGCAGACACACAGCTTTTGGGTAAATTTAA 850
284 hrTyrThrGlnAlaGlnLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
|||||
851 CTTATACTCAAGCAAGAACTTTCAGGCAGCTGCTTAATTCAAAGCAGAGAA 900
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
|||||

901 GGGTTTTTCAACAAGCCATGCTTGAGCAAAACAACAATATCAGTACCAA 950
317 sPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334
|||||
951 ATTTGGTGACATGGCTATTATTATGCTCAAAATCCGCTCTAACAACGCA 1000
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
|||||
1001 TTAATGATGAAGTCGTACGGTTGATGTGGAATATTATTGACCCGTGA 1050
351 HisProValTyrValArgIleAsnPheThrGlyAsnPheLysThrGl 367
|||||
1051 CACCCTGTCTATGTAGCGCGTATTAAATTTACAGTAACCTTTAAAGACCA 1100
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384
|||||
1101 AGATGAAGTACTCGCTGTGAGATGCGACAACCTTGAAGGTGCGTTGGCAT 1150
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
|||||
1151 CTAATCAAAAATCCAGCTGTCTGTCGACGCTTGATCGGACTGGGTTT 1200
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspGl 417
|||||
1201 TTTAAACATGTTACCGTTGATCTCGTCAGTACCCAACTCACCTGATCA 1250
417 nValAspValAsnPheValValGluGluGlnProSerGlySerSerThrI 434
|||||
1251 GGTGTGATGTAAATTTTGGTTGGAAGACAACCTTCAGGATCATCAACCA 1300
434 leAlaAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspVal 450
|||||
1301 TCGCAGCAGGCTACTCTCAAAGTGGTGGTAACTTTTCAATTTGATGTT 1350
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
|||||
1351 TCTCAAAATAACTTTATGGGTACAGGTAAAGCACGTCATGCTGCTTTTC 1400
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
|||||
1401 TCGCTCTGAGACCCGCTGAGGTGTATAGTTTGGGTATGACCAACCCATCT 1450
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
|||||
1451 TTACCGTAAATGGCGTCTCGCAAGCTTGAGTGGCTACTATCGCAAAACC 1500
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGl 517
|||||
1501 AAGTATGATAACAAGAACATTAGTAATATGTACTTGATTCTTATGGTGG 1550
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
|||||
1551 CTCATTAAAGCTATGGATATCCAAATGATGAATAAATCAAGCATAAGCTTG 1600
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
|||||
1601 GTCTGAATGCTGACAATACCAAGCTTCATGGCGGTGCTTTTATGGGCATT 1650
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
|||||
1651 AGTAATGTCAAGCAGCTGATGGCAGATGGTGGCAAAATTCAGTGGATAA 1700
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrAsnAlaIleL 584
|||||
1701 TAAATGGCATTCCTGATTTTAAGCATGATTACACAACCTTACAAATGCCATT 1750
584 euGlyTyrAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
|||||
1751 TGGGGTGGAAATTTCAAGTCTAGATCGCCCTGTATTTTCCAACCAAGGC 1800
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGl 617
|||||

```
1801 ATGAGTCATTCTAGATTGACGGTTGGTTTGGTCATAAAAACTCATCA 1850
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysSerV 634
|||||
1851 AAAAGTGGTTTATCAAGGCAATATCATCGCCATTATCAAAAAATCAG 1900
634 aLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
|||||
1901 TCTTGGTGGATACGCCAAGTAGGTAGCTATGGCAATAATTACCATTTAT 1950
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
1951 GAAAATTTCTATGACGCGCTATGTTGGTTCTGGTTCGTGCTATGATCAATC 2000
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
|||||
2001 CUCITTTGGTCCACGCTCAAGCCCTATTGTGACAGCTCGTGGTCAAC 2050
684 InThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
|||||
2051 AAACACACTAGGAGAGTGTGGTGGTAATGCTTTGGCAACTTTCGGC 2100
701 SerGluLeuLeuLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717
|||||
2101 AGTAGCTGATTTTACCTTTGCCATTTAAAGGTGATTTGGATAGATCAGGT 2150
717 lArgProValIlePheIleGluGlyGlnValPheAspThrThrGlyM 734
|||||
2151 GCGTCAGTGATATTCATTGAGGCGCTCAGGTTTTTGTATACACAGGTA 2200
734 eAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
|||||
2201 TGGATAAACAAACCATTTGATTTAAACCAATTTAAAGACCACCAAGCAACA 2250
751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLy 767
2251 GTGTAACAAATGCAAAAGCAGCCAAATCCCGCTACTAACCCAAAGATAA 2300
767 sGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784
|||||
2301 ACAGTTGCGTTATAGTCTGGTGGTGGTGCACCTTGGTATACGCCCATYG 2350
784 lProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
|||||
2351 GTCCCTTATCTATTAGCTATGCAAGCCATTGAATAAAAAACAAATGAT 2400
801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
2401 CAGACCGATACGGTACAGTTCAGATTGGTAGTGCTTT 2439
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAZ29551

seq_documentation_block:

ID: AAZ29551 standard; DNA; 2442 BP.

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

XX AC AAZ29551;

```
PD 09-DEC-1999.
XX 31-MAY-1999; 99WO-EP03822.
XX 03-JUN-1998; 98GB-0011945.
XX 08-MAR-1999; 99GB-0005304.
XX (SMIK ) SMITHLINE BEECHAM BIOLOGICALS.
XX Vinals-Bassols C;
XX WPI; 2000-105700/09.
XX P-PSDB; AAY44391.
XX Novel BAS027 polynucleotide and polypeptides from Moraxella
XX catarrhalis useful for treating M. catarrhalis infection such as otitis
XX media
XX Claim 8; Page 102; 109pp; English.
XX The present sequence is a DNA obtained from chromosomal DNA library of
XX Moraxella catarrhalis strain Mc2931 (ATCC 43617). It encodes BAS027
XX polypeptide, which shows significant homology to Neisseria meningitidis
XX OMP85 outer membrane protein. BAS027 polynucleotide and polypeptide can
XX be used for diagnosis and staging of disease, determining susceptibility
XX to a disease and to prepare medicaments for treating M. catarrhalis
XX infections, especially otitis media. The BAS027 DNA can be used as
XX probe for screening of genetic mutations, serotype, taxonomic
XX classification or identification. BAS027 agonists, antagonists and
XX antibodies may be used to prevent and/or treat bacterial infections.
XX Sequence 2442 BP; 735 A; 461 C; 535 G; 711 T; 0 other;
```

alignment_scores:

Quality: 4189.00 Length: 813

Ratio: 5.159 Gaps: 0

Percent Similarity: 99.877 Percent Identity: 99.754

alignment_block:

US-09-701-711-2 x AAZ29551 ..

Align seg 1/1 to: AAZ29551 from: 1 to: 2442

```
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetAl 17
|||||
1 ARGCGTAATTCAATATTTAAAGGTTTTTCAAGTCTGAGTCAGTGCATGACATGCC 50
17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaA 34
|||||
51 TGTGTCATGATGGTAAATGCTCAACTCATGTCACAAAGCGCGGATTTTATGGCAA 100
34 sNAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50
|||||
101 ATGACATTGCCATCACAGGACTACAGCGAGTGACCATTTGAAAGCTTACAA 150
51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67
|||||
151 ACGGTGCTCCGCTTCGCTTGGGTCAAGTGGTGAGCGAAGCACAGTTGGC 200
67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84
|||||
201 TGATGGTGTCAAAGCACTTTATGCAACAGGCAATTTTTCAGATGTGCAAG 250
84 aLtyrHisGlnGluGlyArgIleIleIleValThrGluArgProLeu 100
|||||
251 TCTATCATCAAGAAAGGCGTATCATCTATCAGGTAACCCGAAAGGCCGTTA 300
101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117
|||||
301 ATCGCTGACATTAATTTTTCAGGCAATCGCTTAATTCACAAAGAGGTCT 350
117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134
```

|||||
351 ACAGAGGGCTAAAAATGCTGGCTAGCTGTGGTCAACCACTAAAAAC 400
134 lnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150
|||||
401 AAGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATCA 450
151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG1 167
|||||
451 CAAGGCTATTATATACCGAAATCTGTCAACACAGAGATGCTTGATGG 500
167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184
|||||
501 TAATCGTGTAAAGCTTGATATGACCTTTGCTGAAGGTAAACCTGCACGG 550
184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
|||||
551 TGGTTGATATTAAATATCATCTTGGCAATCAGCATTTTATAGCATGCAGATTG 600
201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217
|||||
601 ATTGATGCTTGGCATTAAGGATAATAAATCAATCCACCTGCTAAAGC 650
217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234
|||||
651 TGACCGTTTACTCAAGAAAAGCTGGTGACCCAGTTTACAGAAATTGCGTG 700
234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250
|||||
701 CTAAATATCTCAATCGAGGGTTTGTGCGCTTTTGAGATTAAAGATGCTAAG 750
251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267
|||||
751 CTTAATATTATGAGATTAAGAACCCGTATCTTTGTTGAGATTTCAATGCA 800
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
|||||
801 TGAAGGTGAGCAATATCGCTTGGACAGACACAGTTTTTTGGTAAATTTAA 850
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
|||||
851 CTTATACTCAAGCAACTTGAGGCACCTGCTTAAATTCAGAGCAGAAGAA 900
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
|||||
901 CGGTTTTACAAAGCCATGCTTGAGCAAAACAAACAATATCAGTACCAA 950
317 sPheGlyAspAspGlyTyrTyrTyrAlaGlnIleArgProValThrArgI 334
|||||
951 ATTTGGTGACGATGGCTATTATTATGCTCAAAATCCGCTCTGTAACACGCA 1000
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
|||||
1001 TTAATGATGAAGACTCGTACGGTTGATGGATATATTATGACCCCTGTA 1050
351 HisProValTyrValArgIleAsnPheThrGlyAsnPheLysThrG1 367
|||||
1051 CACCCTGCTATGTACGCGGTATTAATTTACAGGTAACCTTTAAGACCCA 1100
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaS 384
|||||
1101 AGATGAAGTACTCCGCTCGTAGATCGCAACCTTGAAGGTGCGTTGGCAT 1150
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
|||||
1151 CTAATCAAAAAATCCAGCTGCTCGTGACGCTTGATGGGACTGGGTTT 1200
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspG1 417
|||||
1201 TTTAAACATGTTACCGTTGATCTCGTCAGTACCACCACTCACCTGATCA 1250
417 nValAspValAsnPheValValGluGluGlnProSerGlySerSerThrI 434
|||||

1251 GGTGTATGTAATAATTTGTGGTTGAAGAACACCTTCAGGATCATCAACCA 1300
434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
|||||
1301 TCGCAGCAGGCTACTCTCAAAGTGGTGGTAACTTTTCAATTTGATGTT 1350
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
|||||
1351 TCTCAAAATAACTTTTATGGGTACAGGTAAAGCACGCTCAATGCTCGTTTC 1400
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
|||||
1401 TCGCTCTGAGACCGCTGAGGTGTATAGTTTGGGTATGACCAACCCATACT 1450
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
|||||
1451 TTACCGTAATGGGCTCTCGCAAGCTTGAGTGGCTACTACTCGTAAACCC 1500
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyG1 517
|||||
1501 AAGTATGATAACAAGACATTAGTAATTTACTTGTATTCTTATGGTGG 1550
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
|||||
1551 CTCATTAAAGCTATGGATATCAATTTGATGAAAATCAACGCATAAGCTTTG 1600
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
|||||
1601 GTCGAATGCTGACAATACCAGCTTCATGGCGGTCTGTTTATGGGCATT 1650
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
|||||
1651 AGTAATGTCAAGCAGCTGATGGCAGATGGTGCAAAATTCAGTGGATAA 1700
567 nasnGlyIleProaspPheLysHisAspTyrThrThrTyrAsnAlaIleL 584
|||||
1701 TAAATGGCATCTCTGATTTTAAAGCATGATTACACAACCTACAAATGCCATT 1750
584 euGlyTyrAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
|||||
1751 TGGGTGGAATTAATCAAGTCTAGATCGCCCTGTATTTCCAACCCAAGGC 1800
601 MetSerHisSerValAspLeuThrValGlyPheGlyAspLysThrHisG1 617
|||||
1801 ATGAGTCATCTGTAGATTTGACGGTTGGTTTGGTGATAAACTCATCA 1850
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerV 634
|||||
1851 AAAAGTGGTTTATCAAGGCAATATCTATCGCCCATTTATCAAAAATCAG 1900
634 alLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
|||||
1901 TCTTCGCTGGATACGCCAAGTTAGGCTATGGCAATAATTTACCATTTAT 1950
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
1951 GAAAAATTTCTATGACGGCGGTATGGTTCGGTTCGGTATGATCAATC 2000
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||
2001 CTCCTTTGGTCCACGCTCACAAGCCTATTGACAGCTCGTCTGTCACAC 2050
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
|||||
2051 AAACCCACACTAGGAGAGGTTGTTGGTGGTAAATGCTTTGGCAACTTTCGGC 2100
701 SerGluLeuIleLeuProIleProPheLysGlyAspTrpIleAspGlnVa 717
|||||
2101 AGTGAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATAGATCAGGT 2150
717 lArgProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734
|||||
2151 GCGTCCAGTGATATTCAATTGAGGCGGTGAGGTTTGTGATACAACAGGTA 2200

734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
 |||||
 2201 TGGATAAACAACCATTTGATTTAAACCCAAATTTAAAGACCCACACAAACA 2250
 |||||
 751 AlaGluGlnAsnAlaLysAlaAlaAsnArgProLeuLeuThrGlnAspLys 767
 |||||
 2251 GCTGAACAAAATGCAAAAGCAGCCCAATGCCCGCTACTAACCCCAAGATAA 2300
 |||||
 767 scLInLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784
 |||||
 2301 ACAGTGTGGTTATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2350
 |||||
 784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
 |||||
 2351 GTCCCTTATCTATAGTATGCAAGCCATTGAATAAAAACAATAATGAT 2400
 |||||
 801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
 |||||
 2401 CAGACCGATACGGTACAGTTCAGATTCCAGATTGGTAGTGCTTTT 2439
 |||||

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA2001A.DAT:AAF28550

seq_documentation_block:

ID AAF28550 standard; DNA; 99629 BP.

XX AAF28550;

XX 04-APR-2001 (first entry)

XX Genomic fragment #37.

XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.

XX Moraxella catarrhalis.

XX WO200078968-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US16649.

XX 18-JUN-1999; 99US-0140121.

XX (INCY-) INCYTE GENOMICS INC.

XX Lagace RE, Patterson C, Berg KL;

XX WPI; 2001-041427/05.

XX Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -

PS Claim 1; Page 391-415; 545pp; English.

XX The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.

XX Sequence 99629 BP; 29233 A; 19222 C; 21909 G; 29264 T; 1 other;

alignment_scores:

Quality: 4189.00 Length: 813
 Ratio: 5.159 Gaps: 0
 Percent Similarity: 99.877 Percent Identity: 99.754
 alignment_block:
 US-09-701-711-2 x AAF28550 ..
 Align seg 1/1 to: AAF28550 from: 1 to: 99629
 1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetal 17
 |||||
 66133 ATGCGTAATTCATATTTAAAGGTTTTCAGGTGAGTGCATGACATGACATGGC 66182
 |||||
 17 aValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheMetAlaA 34
 |||||
 66183 TGTCTATGATGGTAAATGTCACACTCATGCACACAGCGGGGATTTTATGGCAA 66232
 |||||
 34 snAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGln 50
 |||||
 66233 ATGACATTTGCCATCACAGGACTACAGCGAGTGACCATTTGAAAGCTTACAA 66282
 |||||
 51 SerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAl 67
 |||||
 66283 ACGGTGCTGCCGTTTCGCTTGGGTCAGGTGAGCGAAGCACAGTTGGC 66332
 |||||
 67 aAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnV 84
 |||||
 66333 TGTATGTTCAAGACACTTTATGCAACAGGCAATTTTCAGATGTGCAAG 66382
 |||||
 84 alTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeu 100
 |||||
 66383 TCTATCATCAAGAGGGCGTATCATCTATCAGGTAACCGAAAGGCGGTTA 66432
 |||||
 101 IleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGlyLe 117
 |||||
 66433 ATCGCTGAGATTAAATTTGAGGGCAATCGCTTAATTTCCAAAGAGAGGTCT 66482
 |||||
 117 uGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysG 134
 |||||
 66483 ACAAGAAGGGCTAAAAATGCTGGCTTAGCTTGGGTCAACCCACATAAAC 66532
 |||||
 134 InAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSer 150
 |||||
 66533 AAGCCACAGTACAGATGATCGAAACCGAGCTTACCAATCAATATATATCA 66582
 |||||
 151 GlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspG1 167
 |||||
 66583 CAAGGCTATTATATACCGAAATTTACTGTCAACACAGACCATGCTTGATGG 66632
 |||||
 167 yAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgV 184
 |||||
 66633 TAATCGTGTAAAGCTTGATATGACCTTTGCTCAAGGTAAACCTGCACGGG 66682
 |||||
 184 alValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeu 200
 |||||
 66683 TGGTTGATATTAAATATCATTTGGCAATTCAGCATTTTAGCGATGCAGATTG 66732
 |||||
 201 IleAspValLeuAlaIleLysAspAsnLysIleAsnProLeuSerLysAl 217
 |||||
 66733 ATGTATGCTTGGCTTAAGGATAAATAAATCAATCCACTGCTTAAAGC 66782
 |||||
 217 aAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuArgA 234
 |||||
 66783 TGACCGTTATACTCAAGAAAAGCTGGTGACCACTTTAGAGAAATTTGCGTG 66832
 |||||
 234 laLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLys 250
 |||||
 66833 CTAAATATCTCAATCCAGGGTTTGTGCGCTTTTGAGATTAAAGATGCTAAG 66882
 |||||
 251 LeuAsnIleAsnGluAspLysAsnArgIlePheValGluIleSerLeuHi 267
 |||||
 66883 CTTAATATTAATGAAGATAAAACCGTATCTTTTGTGAGATTTCATTGCA 66932
 |||||

```
267 sGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuT 284
|||||
66933 TGAAGGTGACAATATCGCTTGACAGACACAGCTTTTGGGTAATTAA 66982
284 hrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLysAlaGluGlu 300
|||||
66983 CTTATACTCAAGCAGAAGCTTGAGGCACCTGCTTAAATTCAAAGCAGAAGAA 67032
301 GlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLy 317
|||||
67033 GGGTTTTCACAAAGCCATGCTTGACCAACCAACAAACAATATCATGACACAA 67082
317 sPheGlyAspAspGlyTyrTyrAlaGlnIleArgProValThrArgI 334
|||||
67083 ATTTGGTGACGATGGCTATATATATGCTCAATCCGTCCTGTACACGCA 67132
334 leAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProVal 350
|||||
67133 TTAATGATGAAGTCGTACGGTTGATGTGGAATATTATATTGACCCCTGTA 67182
351 HisProValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrG 367
|||||
67183 CACCCTGTCTATGTACCGCGTATTAAATTTTACAGGTAACTTTTAAGACCCA 67232
367 nAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlas 384
|||||
67233 AGATGAAGTACTCCGTCGTGAGATGCACAACTTTGAAGTGCCTGGCAT 67282
384 erAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhe 400
|||||
67283 CTAATCAAAAAATCAGCTGCTCGTCGACGCTTGATCGGACTGGGTTT 67332
401 PheLysHisValThrValAspThrArgProValProAsnSerProAspG 417
|||||
67333 TTTAAACATGTTACCGTTGATCTGCTCCAGTACCCAACTCACCTGATCA 67382
417 nValAspValAsnPheValValGluGlnProSerGlySerSerThrI 434
|||||
67383 GGTTCATGTAATTTTGTGGTGAAGAACACACCTTCAGGATCATCAACCA 67432
434 leAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspVal 450
|||||
67433 TCGCAGCAGGCTACTCTCAAAGTGGTGGTAACTTTTCAATTTGATGTT 67482
451 SerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSe 467
|||||
67483 TCTCAAAATAACTTTATGGGTACAGGTAAGCACGCTCAATGCTTCGTTTC 67532
467 rArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrP 484
|||||
67533 TCGCTCTGAGACCCGTCGAGGTGTATAGTTTGGGTATGACCAACCCACT 67582
484 heThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThr 500
|||||
67583 TTACCGTAAATGGGCTCTCGCAAGCTTGAGTGGCTACTATCGTAAACCC 67632
501 LysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyG 517
|||||
67633 AAGTATGATAACAAGAACAATAGTAATATGTAATGTAATCTTATGTTGG 67682
517 ySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheG 534
|||||
67683 CTCATTAGCTTATGGATATCAATGGAATGATGAAAATCAACGCATAAGCTTG 67732
534 lyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIle 550
|||||
67733 GTCTGAATGCTGACAATACCAAGCTTCATGCGCGCTTTTATGGGCATT 67782
551 SerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAs 567
|||||
67783 AGTAATGTCAAGCAGCTGATGGCAGATGGTGCAAAATTCAGGTGCGATAA 67832
567 nAsnGlyIleProAspPheLysHisAspTyrThrThrAsnAlaIleL 584
```

```
|||||
67833 TAATGGCATTCCTGATTTTAAGCATGATTACACAACCTACAATGCCATTT 67882
584 euGlyTyrAsnTyrSerSerLeuAspArgProValPheProThrGlnGly 600
|||||
67883 TGGGGTGGAAATTAATCAAGTCTAGATCGCCTGTATTTCCAACCCAAGGC 67932
601 MetSerHisSerValAspLeuThrValIclyPheGlyAspLysThrHisG 617
|||||
67933 ATGAGTCATTTCTGTAGATTTGACGGTTGGTTTGGTGATAAAACTCATCA 67982
617 nLysValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerV 634
|||||
67983 AAAAGTGGTTTATCAAGGCAATATCTATCGCCCATTTATCAAAAAATCAG 68032
634 alLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyr 650
|||||
68033 TCTTGGCTGGATACGCCAAGTTAGGCTATGGCAATAATTTACCATTTTAT 68082
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
68083 GAAATTTCTATGCAGGGGCTATGGTTCGGTTCGTGCTATGATCAATC 68132
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||
68133 CTCCTTTGGTCCACGCTCACAAAGCCTATTTGACAGCTCGTCGTGTCACAC 68182
684 lnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheGly 700
|||||
68183 AAACCACTAGGAGAGGTTTGGTGGTAAATGCTTTGGCAACTTTTCGGC 68232
701 SerGluLeuIleLeuProLeuProPheLysGlyAspTrpIleAspGlnVa 717
|||||
68233 AGTCAGCTGATTTTACCTTTGCCATTTAAAGGTGATTGGATAGATCAGGT 68282
717 lArgProValIlePheIleGluGlyGlyGlnValPheAspThrThrGlyM 734
|||||
68283 GCGTCCAGTGTATTCATTTAGGCGCGTCAAGTTTGGTATACAAAGGTA 68332
734 etAspLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThr 750
|||||
68333 TGGATAACAAACCATTTGATTTAACCCCAATTTAAAGACCCCAAGCAACA 68382
751 AlaGluGlnAsnAlaLysAlaAsnArgProLeuLeuThrGlnAspLys 767
|||||
68383 GCTGAACAAAATGCAAAAGCAGCAATTCGCCGCTACTAACCCCAAGATAA 68432
767 sGlnLeuArgTyrSerAlaGlyValGlyAlaThrTrpTyrThrProIleG 784
|||||
68433 ACAGTTGCGTTATAGTCTGCTGGTGGTGCAACTTGGTATACGCCCATTTG 68482
784 lyProLeuSerIleSerTyrAlaLysProLeuAsnLysLysGlnAsnAsp 800
|||||
68483 GTCCTTTATCTATTAGTATGCCAAGCATTGAATAAAAAACAATAATGAT 68532
801 GlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
|||||
68533 CAGACCATGAGGTACAGTTCCAGATTGGTAGTGTCTTT 68571
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAS07278
seq_documentation_block:
ID AAS07278 standard; DNA; 2379 BP.
XX
AC AAS07278;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup A antigenic protein DNA.
XX
KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
bacterial infection; baculovirus; yeast; ds.
XX
```

OS Neisseria meningitidis.

XX Key Location/Qualifiers

XX 1..2379

XX CDS

XX /tag= a

XX /product= "N. meningitidis serogroup A antigen"

XX sig_peptide

XX 1..63

XX /tag= b

XX mat_peptide

XX 64..2376

XX /tag= c

XX /product= "Mature N. meningitidis serogroup A antigen"

XX WO200138350-A2.

XX 31-MAY-2001.

XX 28-NOV-2000; 2000WO-IB01851.

XX 29-NOV-1999; 99GB-0028197.

XX 09-MAR-2000; 2000GB-0005698.

XX (CHIR-) CHIRON SPA.

XX (STAT-) STATENS INST FOLKEHELSE.

XX Giuliani MM, Pizza M, Rappuoli R, Holst J;

XX WPI; 2001-381289/40.

XX P-PSDB; AAU03958.

XX Novel 85 kDa antigen from Neisseria meningitidis and Neisseria gonorrhoeae, useful in the manufacture of a medicament for treating and preventing Neisserial bacteria infection -

XX Claim 6; Fig 2; 92pp; English.

XX The sequence represents a DNA encoding a Neisseria meningitidis serogroup A 85 kDa antigenic protein. Neisseria meningitidis colonises the pharynx, causing meningitis and, occasionally, septicaemia in the absence of meningitis. This antigenic protein is useful in the manufacture of a medicament for treating or preventing infection due to Neisseria bacteria, such as meningitis and septicaemia. It is also useful as a diagnostic reagent for detecting the presence of Neisseria bacteria or antibodies raised against Neisseria, and as a reagent for raising the variety of different expression systems, for example, mammalian cells, baculoviruses, plants, bacteria and yeast.

XX Sequence 2379 BP; 636 A; 755 C; 596 G; 392 T; 0 other;

alignment_scores:

Quality: 1123.50 Length: 837

Ratio: 2.073 Gaps: 19

Percent Similarity: 64.755 Percent Identity: 32.616

alignment_block:

US-09-701-711-2 x AAS07278 ..

Align seg 1/1 to: AAS07278 from: 1 to: 2379

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26

4 AAACCTGAACAGATTGCCCTCCGCACTGATGATGTTGGGCATATCGGCTTT 53

26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrGlyLeuGlnA 43

54 GGCAATTGCCGACTTACCATCCAGACATCCGTGTGGAAGGCTTGACG 103

43 rgValThrIleGlnSerLeuGlnSerValLeuPropPheArgLeuGlyGln 59

104 GTACCGAGCGGAGCAGCGATTTCACACTACCTGCCGTCACAAAGTCGGCGAC 153

60 ValValSerGluAsnGlnLeuAlaAlaAspGlyValValLysAlaLeuTyrAla 76

154 ACCTACACGACACACACGCGAGTGCATCATCAAAAGCCTGTACGCCAC 203

76 rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIle 93

204 CGGTTCTTTGAGGAGCTAGAGTCAAACTGCGGAGCGGCGAGCTCTGC 253

93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109

254 TGACCGTTATTCGAACGCCCAACCATCGGCTCGCTCAACATCACCGCGCC 303

110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126

304 AAAATGCTCAAAACACGCCCATCAAGAAAAACCTCGAATCGTTGGGCT 353

126 uAlaValGlyGlnProLysGlnAlaThrValGlnMetIleGluThrG 143

354 GGGCGAGTCGCAATACTTTAATCAGGCGACACTCAACAGGCGAGTCGCCG 403

143 luLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr 159

404 GCCTGAAGAAGAATAACCTCGGCGTGGCAAACTCAATATCCAAATCACG 453

160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176

454 CCCAAAGTAACCAAACTCGCCCAACCGCTCGACATCGACATCAGCAT 503

176 eAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG 193

504 TGACGAGGGCAATCCGCCAAATCACCGACATCGAATTTGAAGGCAACC 553

193 lnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209

554 AAGTCTATTCCGACCGCAAACTGATCGCGCAGATGTCGCTGACCGAAGGC 603

210 LysIle...AsnProLysSerLysAlaAspArgTyrThrGlnGluLysLe 225

604 GGCATTGTGACATGGTGTACAGAGCAGCCGGTTCGACCGCGCAGAAAT 653

225 uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV 242

654 CGCCCAAGACATGGAAGAAAGTAAACCACTTCTACAGAGCAACCGGTACT 703

242 alArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258

704 TCGATTTCGTCATCCTCGATACCGACATCCAAACCAACGAAGACAAAACC 753

259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheG 275

754 AGGCAGACCATCAAAATCACCGTCCACGAAGCGGACGTTTCGCGCTGGGG 803

275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG 291

804 CAAAGTGTGATTGAAGCGCACACCAACGAAGTCCCAAGCGCGCAACTGG 853

291 luAlaLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeu 307

854 AAAAAGTGTGCTGACCATGAAGCGCGCAAAATGGTACGAACGCCAGCATG 903

308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyrTy 324

904 ACCGCGGTTTTGGGTGAGATTCAAGAACCGCATGGGCTCGGCGAGGTACGC 953

324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341

954 ATACACGGAAATCAGCGTACGCGCTCGCGGACCGCGGCAACCAAAACCG 1003

341 alAspValGluTyrTyrIleAspProValHisProValTyrValArgArg 357

1004 TCGATTTCGTCCTCGACATCGAACCGCGCGGAAATCTACGTCAACGAA 1053

358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgG 374

[illegible]

seq_documentation_block:

XX
AC AAH42129:

DE Nucleotide sequence of a *Neisseria gonorrhoeae* protein.

QS *Neisseria gonorrhoeae*.

FT 1..2379 /tag= a /product= "Neisseria serogroup B protein"

```
FT /*tag= b
```

```
FT
...
/*tag= c
```

XX
NA
-C887CT0070M

26-JUL-2001.


```

424 lGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlns 441
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1254 GACGAACGCTCCACCGGCTCGCTGACCTGAGCGCGGCTGGTTTCAGG 1303
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPhMetGly 457
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1304 ATACCGGCTTGGTTCATCGCGCGGCTATCGCAGGACACCTGTTCCGTT 1353
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGlnThrArgGluVa 474
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1354 ACGGCAAGTCGGCGGCTCGCGGCTCGCGAAGCAAAACACCGCTCAA 1403
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1404 CGGCTCGCTGCTGTTACCGACCGTACTTCACGGCAGCGGGGTACG. 1452
491 lnsrLeuSerGlyTyr.....TyrArgLysThr.....LysTyr 502
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1453 .....CTGGGCTAGATATTTACGAAAGACCTTCGACCCCGCAAA 1494
503 AspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLe 519
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1495 GCATCGACCGCGTCAAAACATATAAAACACCAACCGCGCGGCGCGT 1544
519 userTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1545 AAGGATGGTATCCCGCTTACCGAATACGACCGGTCATTTTCGGGCTGG 1594
536 snAlaAspAsnThrLysLeuHisGly..... 544
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1595 CGCGGAACACCTGACCGCTCAACACCTACAAAGCAACCAACGCTAT 1644
545 .....GlyArgPheMetGlyIleSerAsnVally 554
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1645 GCCGACTTTATCAGGAATACGGCAAAACCGACGCG..... 1680
554 sGlnLeuMetAlaAspGlyGlyIleGlnValAspAsnAsnGlyIleP 571
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1681 .....GCACGCGC..... 1689
571 roAspPheLysHisAspTyrThrTyrAsnAlaIleLeuGlyTyrPAsn 587
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1690 ..AGCTTCAAAAGC.....CTGCTGTACAAAGGCACCGCTCGGCTGGGG 1731
588 TyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHisSe 604
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1732 CGCAACAGACCGACGCGCTATGCGCGGACGCGGCTACCTGACCGG 1781
604 rValAspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysV 619
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1782 CGTAATATCCGAATCGCCCTGCCCGCGCAGCAAACTGCAATACTACTCCG 1831
619 alValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeu 635
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1832 CCACCCACAAACCTGTGTTCTTCCTCCCTTAAGCAAAACCTTCACGCTG 1881
636 Arg.....GlyTyrAlaLysLeuGlyTyrGly.....As 645
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1882 ATGCTCGCGCGGAGTCGGCATTCGGGCG...GGCTACGGCAGCAACAA 1928
645 nAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValA 662
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1929 AGAAATCCCTTCTTTGAAAACTTCTACGGCGGCGGCTGGGTTTCGGTGC 1978
662 rgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThr 678
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
1979 GCGGCTACGAAAGCGGACGCTCGCGCGGAA..... 2010
679 AlaArgArgGlyGlnGlnThrThrLeuGlyGlnValVal.....GlyG 693
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2011 .....GTGTATGACCAATACGGCGGCAAAATATCAGCTACGCGGG 2048

```

```

693 yAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheL 710
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2049 CAACAAAAAGCAACCTCTCCGCGAGCTGCTCTCCGATGCCGGTG 2098
710 yGlyAspTrpIleAspGlnValArgProValIlePheIleGluGlyGly 726
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2099 CGAAAGAC...GCACGACCGCTCGCTGAGCCTGTTTGGCGACGAGGC 2145
727 GlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrG 743
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2146 AGCGTGCG.....GACGGCAGAACCTATACCGCGCGCGA 2180
743 nPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnA 760
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2181 AAGCGGTAACAACAATCGGTTTACTCGGAAACGCGCATAAATCC.... 2236
760 rgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGly 776
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2227 .....ACCTTTACCAACAATTCGGCTATTCGCGCGCGCGCGCG 2265
777 AlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysPr 793
: |||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2266 GTTACCTGGCTCTCGCCTTTGGGTCCGATGAAATTCAGTACGCTACCC 2315
793 oLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleG 810
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2316 GCTGAGAAAAAACCGGAAGACGAAATCCACGCTTCCAATTCAGCTCG 2365
810 lySerValPhe 813
|||||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||::: :|||:::
2366 GCACGACGTTTC 2376

seq_name: /SIDS1/gcgdata/geneseq/geneseq-n-emb1/NA2001A.DAT:AAS07277
seq_documentation_block:
ID AAS07277 standard; DNA; 2394 BP.
XX
AC AAS07277;
XX
DT 23-OCT-2001 (first entry)
XX
DE Neisseria meningitidis serogroup B antigenic protein DNA.
XX
KW Serogroup B antigen; pharynx; meningitis; septicaemia; mammalian cell;
KW bacterial infection; baculovirus; yeast; ds.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
CDS 1..2394
FT /*tag= a
FT /product= "N. meningitidis serogroup B antigen"
FT sig_peptide 1..63
FT /*tag= b
FT mat_peptide 64..2391
FT /*tag= c
FT /product= "Mature N. meningitidis serogroup B antigen"
XX
PN WO200138350-A2.
XX
PD 31-MAY-2001.
XX
PF 28-NOV-2000; 2000WO-IB01851.
XX
PR 29-NOV-1999; 99GB-0028197.
PR 09-MAR-2000; 2000GB-0005698.
XX
PA (CHIR-) CHIRON SPA.
PA (STAT-) STATENS INST FOLKEHELSE.
XX
PI Giuliani MM, Pizza M, Rappuoli R, Holst J;
XX
WPI; 2001-381289/40.
DR

```



```
1404 CCGCTCGCTGCTGTTACTACCGCTACTTACGGCAGACGGGGTCAGCC 1453
491 InSerLeuSerGlyTyrTyrArg.....LysThrLysTyrAspAsnLys 505
1454 TGGGCTACGATGTTTACGGAAGAGCCCTCGACCCCGCGCAAGCATCGACC 1503
506 AsnLeuSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrG1 522
1504 AGCATCAACAATATAAAACACCAACCGCAGCGCAGGCATCCGCATGAG 1553
522 yTyrProLeuAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
1554 CGTCCCTGTTACCGAATACGACCGCTGTAATTTCGTTTGGTGGCAGAAC 1603
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
1604 ACCTGACCGCTCAAC.....ACCTACAACAAGCCGCCCAAA 1638
556 LeuMetAlaAspGlyGlyIleGlnValAspAsnAsnGlyIleProAs 572
1639 CACTATGCGGACTTATCAAGAATAACGGCAAAACCGACGGC...ACAGA 1685
572 pPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrS 589
1686 CGCAGAGCTTCAAGAGCTGGCTGTACAAGGTACCGTGGCTGGGGCGCA 1735
589 erSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
1736 ACAAAACCGCAGCGGTATATGGCCGACGGCCGGCTACCTACGCGGGGTG 1785
606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620
1786 AACGCCGAATCGCCCTGCTGCGGCAAACTGCATACTACTCGGCCAC 1835
620 lTyrglnGlyAsnIleTyrArgProPheIleLysLysSerValLeuArg. 636
1836 CCACAACCAACACCTGGTCTTCCCGCTGAGCAAAACCTTACGCTGATGC 1885
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
1886 TCGCGGGGAAGTCGGATTCGGGCG...GGCTACGGCAGAACCAAGAA 1932
647 LeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgG1 663
1933 ATCCCTCTTTTGAAACTTCTACGGCGCGCTGGTTCGGTTCGGTTCGGCG 1982
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaA 680
1983 ATACGAAAGCGGCACGCTCGGTCCGAAA..... 2010
680 rgArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
2011 .....GTCTATGACCAATACCGCGCAAAATAACAGTCAGCGCGGCAAC 2052
695 AlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLysG1 711
2053 AAAAAAGCAACAGCTCTCGCGGAGTGTCTTCCCGATGCCCGCGCGAA 2102
711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlnV 728
2103 AGAC...CGCGGCGACCGTCCGCTGTGGCTGTGGCGCAGCGCAGCG 2149
728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
2150 TGTGGGACGGCAAAACCTACGACCAACACAGCAGTTCGGCGACC..... 2193
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
2194 .....GGCGGCGAGGTTCAAAACATTTACGCGCGCGCAATACCCA 2234
761 oLeuLeuThrGlnAspGlnLeuArgTyrSerAlaGlyValGlyAlat 778
```

```
2235 TAAATCCACCTTTACCAACGAATTGCGCTATTCGCCGCGCGCGGGTTA 2284
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
2285 CCTGGCTCTCGCCTTAGGCCCGCATGAATTCAGTACGCTACCGCTG 2334
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlyse 811
2335 AAGAAAAACCGGAAGCAAGAAATCCACGCTTCCAATTCACCACTCGGCAC 2384
811 rValphe 813
2385 GACGTTT 2391
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH42128
seq_documentation_block:
ID AAH42128 standard; DNA; 2394 BP.
XX
AC AAH42128;
XX
DT 17-SEP-2001 (first entry)
XX
DE Nucleotide sequence of a Neisseria serogroup B protein.
XX
KW Serogroup B protein; outer membrane protein; Neisserial infection;
KW vaccine; ss.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
CDS 1..2394
FT /*tag= a
FT /*product= "Neisseria serogroup B protein"
FT sig_peptide 1..63
FT /*tag= b
FT mat_peptide 64..2391
FT /*tag= c
XX
PN WO200152885-A1.
XX
PD 26-JUL-2001.
XX
PF 17-JAN-2001; 2001WO-IB00166.
XX
PR 17-JAN-2000; 2000GB-0001067.
PR 09-MAR-2000; 2000GB-0005699.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Pizza M, Rappuoli R, Giuliani M;
XX
WPI: 2001-451895/48.
DR P-PSDB; AAB84744.
XX
Composition for treating or preventing infection to, detecting, or for
raising antibodies against Neisserial bacteria, comprises an N.
meningitidis serogroup B outer membrane preparation and an immunogenic
component -
XX
Disclosure; Page 58-59; 83pp; English.
XX
The present sequence encodes a Neisseria serogroup B protein. The
protein is used to produce the compositions of the invention. The
specification describes a composition, comprising a Neisseria
meningitidis serogroup B outer membrane preparation and an immunogenic
component. The immunogenic component is protein disclosed in WO99/57280,
WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
TbpA, TbpB, PilC, Opa, or Omp85. The composition is used for making
a medicament for treating or preventing infection due to Neisserial
bacteria; a diagnostic reagent for detecting the presence of Neisserial
bacteria or of antibodies raised against Neisserial bacteria; and/or
```

CC a reagent which can raise antibodies against Neisserial bacteria. It may
 CC also be used as a vaccine.

XX
 50 Sequence 2394 BP; 657 A; 743 C; 583 G; 411 T; 0 other;

alignment_scores:
 Quality: 1115.00 Length: 819
 Ratio: 2.042 Gaps: 14
 Percent Similarity: 66.667 Percent Identity: 32.479

alignment_block:
 US-09-701-711-2 x AAH42128 ..

Align seg 1/1 to: AAH42128 from: 1 to: 2394

```

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26
   :::::::::::::: :::::::::::::: :::::::::::::: ::
4  AAACGTAAACAGATTGCTTCCGCACGTGATGTTGGGCATATCGCCTTT 53
   :::::::::::::: :::::::::::::: :::::::::::::: ::
26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43
   :::::::::::::: :::::::::::::: :::::::::::::: ::
54 GGCACCTGCGGACTTCACCATCCAGACATCCGCGTGCAGAGCTTGCAGC 103
   :::::::::::::: :::::::::::::: :::::::::::::: ::
43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGln 59
   :::::::::::::: :::::::::::::: :::::::::::::: ::
104 GTACGAGCGGAGTACCGTATTCACTACCTGCCGTCAAAGTCGGCGAC 153
   :::::::::::::: :::::::::::::: :::::::::::::: ::
60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyraLath 76
   :::::::::::::: :::::::::::::: :::::::::::::: ::
154 ACCTCAACGACACACAGCGGAGTGCCTATCATCAAAAGCCTGTACGCCAC 203
   :::::::::::::: :::::::::::::: :::::::::::::: ::
76 rGlyAsnPheSerAspValGlnValTyrrHisGlnGluGlyArgIleIle 93
   :::::::::::::: :::::::::::::: :::::::::::::: ::
204 CGGTTTCITTGACGAGTACGCGTGCAGAACTCGGACGGCGGACCTCTGC 253
   :::::::::::::: :::::::::::::: :::::::::::::: ::
93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109
   :::::::::::::: :::::::::::::: :::::::::::::: ::
254 TGACCGGTTATCGAAGCCCAACCGCTGCTCAACATCATCACCAGCGCA 303
   :::::::::::::: :::::::::::::: :::::::::::::: ::
110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126
   :::::::::::::: :::::::::::::: :::::::::::::: ::
304 AAATGCTGCAAAAGACGCCCACTTAAAGAAACCTCGAATCGTTCGGGCT 353
   :::::::::::::: :::::::::::::: :::::::::::::: ::
126 uAlaValGlnProLysGlnAlaThrValGlnMetIleGluThrG 143
   :::::::::::::: :::::::::::::: :::::::::::::: ::
354 GCGCGAGTCGCAATACTTTAATCAGCGACACTCAATCAGGCACTCGCG 403
   :::::::::::::: :::::::::::::: :::::::::::::: ::
143 luLeuThrAsnGlnTyrrIleSerGlnGlyTyrrAsnThrGluIleThr 159
   :::::::::::::: :::::::::::::: :::::::::::::: ::
404 GCCTGAAAGAGAATACCTCGGGCGGCAAACTCAATATCCAAATCAGC 453
   :::::::::::::: :::::::::::::: :::::::::::::: ::
160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176
   :::::::::::::: :::::::::::::: :::::::::::::: ::
454 CCCAAGTAACCAAACTCGCCGCAACGCGTGCACATCGACATCAGCAT 503
   :::::::::::::: :::::::::::::: :::::::::::::: ::
176 eAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG 193
   :::::::::::::: :::::::::::::: :::::::::::::: ::
504 TGACGAGGCAAAATCCGCCAAATCACCACATCGAATTTGAAGCAACC 553
   :::::::::::::: :::::::::::::: :::::::::::::: ::
193 lnhIlePheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209
   :::::::::::::: :::::::::::::: :::::::::::::: ::
554 AAGTCTATTCCGACGCAAACTGATCGGCAAAATGTCCTGTACCGAAGGC 603
   :::::::::::::: :::::::::::::: :::::::::::::: ::
210 LysIle...AsnProLeuSerLysAlaAspArgTyrrThrGlnGluLysLe 225
   :::::::::::::: :::::::::::::: :::::::::::::: ::
604 GGCATTGTGACATGCTGACAGAGCAACCAATTCACAGCAGCAAAAT 653
   :::::::::::::: :::::::::::::: :::::::::::::: ::
225 uValThrSerLeuGluAsnLeuArgAlaLysTyrrLeuAsnAlaGlyPheV 242
   :::::::::::::: :::::::::::::: :::::::::::::: ::
654 TGCCCAAGATATGGAAGAAGTACCGACTTCTACCAAAATACGGCTACT 703
   :::::::::::::: :::::::::::::: :::::::::::::: ::
242 alArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258

```

```

704 TCGATTTCGTATCTCTCGATACCGACATCCAAACCAACGACGACAAACC 753
   :::::::::::::: :::::::::::::: :::::::::::::: ::
259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrrArgPheGl 275
   :::::::::::::: :::::::::::::: :::::::::::::: ::
754 AAGCAGACCATCAAAATCACCCTCCACGAAGCGGACGTTTCGTTGGGG 803
   :::::::::::::: :::::::::::::: :::::::::::::: ::
275 yGlnThrGlnPheLeuGlyAsnLeuThr...TyrrThrGlnAlaGluLeuG 291
   :::::::::::::: :::::::::::::: :::::::::::::: ::
804 CAAAGTCTCCATCGAAGCGCACACCAACGAAGTCCCCAAAGCCGAACCTG 853
   :::::::::::::: :::::::::::::: :::::::::::::: ::
291 luAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeu 307
   :::::::::::::: :::::::::::::: :::::::::::::: ::
854 AAAAAGTCTGCTGACCATGAAGCCCGCAATATGGTACGAACCCAGCAGATG 903
   :::::::::::::: :::::::::::::: :::::::::::::: ::
308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyrrTy 324
   :::::::::::::: :::::::::::::: :::::::::::::: ::
904 ACCGCGGTTTGGTGAGATTGAGAACCCGATGGGCTCGGACAGGTACGC 953
   :::::::::::::: :::::::::::::: :::::::::::::: ::
324 rTyrrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
   :::::::::::::: :::::::::::::: :::::::::::::: ::
954 ATACAGCGAATCAGCGTACAGCGCTGCCGACGCTGAACCAACCAACCG 1003
   :::::::::::::: :::::::::::::: :::::::::::::: ::
341 alaspValGluTyrrTyrrIleAspProValHisProValTyrrValArgArg 357
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1004 TCGATTTCGTCTCTGCACATCGAACCGGCGGAAATCTACGTCAACGAA 1053
   :::::::::::::: :::::::::::::: :::::::::::::: ::
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgG 374
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1054 ATACACATCACCAGGCAACAAACCCGCGAGAGTCTCGCGCGTGA 1103
   :::::::::::::: :::::::::::::: :::::::::::::: ::
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1104 ATTACGCCAAATGGATCCGCACCTTACGACACCTCCAAAGCTCAACGTT 1153
   :::::::::::::: :::::::::::::: :::::::::::::: ::
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValasp 407
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1154 CCAAGAGCGCGCTCGAGCTTTTGGGTACTTCGACAATGTCCAGTTTGAT 1203
   :::::::::::::: :::::::::::::: :::::::::::::: ::
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1204 GCTGTCCCGCTTGGCGGACGCCGCGCAAGTCAAGTCAATGAACATGAGTCT 1253
   :::::::::::::: :::::::::::::: :::::::::::::: ::
424 lGluGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrrSerGlnS 441
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1254 GACCGAAGCTTCCACCGGTTCCCTGGATTGAGCGGGTGGGTTCAAG 1303
   :::::::::::::: :::::::::::::: :::::::::::::: ::
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1304 ATACCGGGTTGGTCATGTCGCGAGGCGTTTCCCAAGCAACCTGTTCGGT 1353
   :::::::::::::: :::::::::::::: :::::::::::::: ::
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1354 ACGGCAAGTCCGCGCACCTGCGCGCTCCAGAGCAACCAACCGCTTAA 1403
   :::::::::::::: :::::::::::::: :::::::::::::: ::
474 lTyrrSerLeuGlyMetThrAsnProTyrrPheThrValAsnGlyValSerG 491
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1404 CGGCTCGCTGCTGTTTACTGACCGTACTTACGGCAGAGCGGGTTCAGCC 1453
   :::::::::::::: :::::::::::::: :::::::::::::: ::
491 lnsrLeuSerGlyTyrrTyrrArg.....LysThrLysTyrrAspAsnLys 505
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1454 TGGGTACGATGTTTACGGAAAAAGCTTCGACCCCGCAAGCATCGACC 1503
   :::::::::::::: :::::::::::::: :::::::::::::: ::
506 AsnIleSerAsnTyrrValLeuAspSerTyrrGlyGlySerLeuSerTyrrG 522
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1504 AGCATCAAAACAATATAAAACCCAGCGCGCGCGGATCCGCGATGAG 1553
   :::::::::::::: :::::::::::::: :::::::::::::: ::
522 yTyrrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
   :::::::::::::: :::::::::::::: :::::::::::::: ::
1554 CGTGCTGTACCGAATACGACCGCGTGATTCGGTTTGGTGGCGAGAAC 1603
   :::::::::::::: :::::::::::::: :::::::::::::: ::
539 snThrLysLeuHisGlyGlyArgPheMetGlyLysSerAsnValLysGln 555
   :::::::::::::: :::::::::::::: :::::::::::::: ::

```

```
1604 ACCTGACCGCTAAC.....ACCTACAAACAAAGCCCAAA 1638
556 LeuMetAlaaspGlyGlyLysIleGlnValaspAsnAsnGlyIleProAs 572
1639 CACTATCGCGACTTATCAGAAATACGGCAAAACCGACGCG....ACAGA 1685
572 pPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrPasnTyrS 589
1686 CGGCAGCTTCAAAGCGTGGTGTACAAAGGTACCGTGGTGGGGCGCA 1735
589 erSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
1736 ACAAACCGACAGCGCGCTTATGCGCGACGCGCGCTACCTACGACGGGTG 1785
606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620
1786 AACCGCAAAATCGCCCTGCGTGGCAGCAAACTGCAATACTACTCCGCCAC 1835
620 lTyrGlnGlyAsnIleTyrArgPropheIleLysLysSerValLeuArg. 636
1836 CCACACCAAAACCTGGTCTTCCCTCGAGCAAAACCTTCACGCTGATGC 1885
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
1886 TCGCGCGCGAAGTCGCGATTGCGGCG...GGCTACGGCAGCAACCAAGAA 1932
647 LeuProPheTyrGluAsnPheTyrAlaGlyTyrGlySerValArgG1 663
1933 ATCCCTTCCTTGAATACTTCTACGGCGCGCGCTGGGTTCCGGTGGCGG 1982
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaA 680
1983 ATACGAAGCGGACGCGCTCGCTCGGAAA..... 2010
680 rGArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
2011 .....GTCTATGACGAATACGGCGCAAAATCAGCTACGGCGGCAAC 2052
695 AlaLeuAlaThrPheGlySerGluLeuLeuProLeuProPheLysG1 711
2053 AAAAAGCAACCAACGCTCCGCGAGCTGCTCTCCGATGCGCGCGCGAA 2102
711 YAspTrpIleAspGlnValArgProValIlePheIleGluGlyClnV 728
2103 AGAC...GCCGGCACCGTCCGCTGACCCCTGTTCGCGACGAGCGAGCG 2149
728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
2150 TGTGGGACGCAAAACCTACGACGACACACAGCTTCCGCGGACC..... 2193
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
2194 .....GGCGGAGGGTTCAAAACATTTACGGCGCGCGCAATACCCA 2234
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
2235 TAATCCACTTTACCAACGAATTCGCTATTCGCGCGCGCGCGGTTA 2284
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
2285 CTGGCTCTCGCTTTAGGCGCGATGAATTCAGCTACGCTACCGCTCGGCTG 2334
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
2335 AAGAAAAACCGGAAGACGAATTCACGCTTCCAAATTCCAACTCGGCAC 2384
811 rValPhe 813
2385 GAGCTTC 2391
```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:AAA81478
seq_documentation_block:

```
ID AAA81478 standard; DNA; 52253 BP.
XX
AC AAA81478;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_26 SEQ ID NO:26.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappuoli R, Pizza M;
XX
DR WPI; 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea.
XX
PS Claim 7; Page 532-547; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to
CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against Meningococcus B; against all serotypes;
CC and/or against all pathogenic Neisseriae. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
XX
SQ Sequence 52253 BP; 12877 A; 15510 C; 13103 G; 10761 T; 2 other;
```

```
alignment_scores:
  Quality: 1115.00      Length: 819
  Ratio: 2.042         Gaps: 14
  Percent Similarity: 66.667   Percent Identity: 32.479

alignment_block:
US-09-701-711-2 x AAA81478
```

Align seg 1/1 to: AAA81478 from: 1 to: 52553

10	GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl	26
47301	AAACTGAAACAGATGCTTCGCGACTGATGATGTTGGCATATCGCCTTT	47350
26	aGlnAlaAlaAspMetMetAlaAsnAspIleThrIleThrGlyLeuGlnA	43
47351	GGCAGCTTGGCGACTTCACCATCAAGACATCCGGCTCGAAGCGCTGCAGC	47400
43	rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln	59
47401	GTACCGAGCGAGTACCGTATTCATACCTACCTGCCGTCNAAGTCGGGAC	47450
60	valValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh	76
47451	ACCTACAACGACACACACGCGAGTGCCTCATCAAAAGCCGTGTACGCCAC	47500
76	rgLysAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIle	93
47501	CGGTTCCTTTGACGACGTACGCGTCTGAAACTGCGGACGGCGAGCTCTGC	47550
93	yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn	109
47551	TGACCGTTATCGAAGCGCCACCATCGGCTCGCTCAACATCACCGCGGCA	47600
110	ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe	126
47601	AAAAAGTGTCAAAACGACGCGCATTAAGAAAAACCTCGAATCGTTCCGGCT	47650
126	uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG	143
47651	GCGCGAGTCGCATACTTTAATCAGGCGACACTCAATCAGCGACGTCCGCG	47700
143	luLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr	159
47701	GCCTGAAGAAAGAAATACCTCGGCGCGGCAAACTCAATATCCAATCACG	47750
160	valLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh	176
47751	CCCAAAAGTACCAAACTCGCGCGCAACCGGCTCGACATCGACATCAGAT	47800
176	eAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG	193
47801	TGACGAGGGCAATCCGCGCAAAATCACCGACATCGAATTTGAAGCGAAC	47850
193	lnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn	209
47851	AAGTCATATCCGACCGCAAACTGATCGGCGCAAAATGTCCTGACCGAAGG	47900
210	LysIle...AsnProLeuSerLysAlaAspArgTyrThrGlnGluLysLe	225
47901	GGCATTTGGACATGTGTCACGACGAAGCAACCAATTCAACGACGACAGAA	47950
225	uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV	242
47951	TGCCCCAAGATATGGAAGAAAGTAACCGACTTCTACCAAAAATACCGCTACT	48000
242	aArgPheGluIleLysAlaLysLeuAsnIleAsnGluAspLysAsn	258
48001	TCGATTTCCGTATCCGTATCCGACATCCAAACCAACGAAGACAAACACC	48050
259	ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheGl	275
48051	AAGCAGACCATCAAAATCACCGTCCAGAAAGGGGAGCGTTTCCGTTGGG	48100
275	yGlnThrGlnPheLeuGlyAsnLeuThr...TyrThrGlnAlaGluLeuG	291
48101	CAAAAGTCTCCATCGAAGGCGACCAACGAAGTCCCCCAAGACCGAATGG	48150
291	luAlaLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeu	307

48151 AAAAACTGCTGACCATGAAGCCCGCAATGGTACGAACGCCAGCATG 48200

308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyrTy 324
::: :::::||||:||||:||||:||||:||||:||||: |||||

48201 ACCGCGGTTTTGGGTGAGATTACGAACCGCATGGCTCGGCAGCTACGC 48250

324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
||||:||||:|||| ||| |||:||||:|||| |||

48251 ATPACAGCAAAATCATCGCTACAGCGCTCCGCAAGCGCTGAACACCAACCG 48300

341 aLysValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
||||| ||||| :::::||||:||||: ||::: |||||:||||: ||::: |||||

48301 TCGATTTCGTCTGCATCATCAAGACGGCGCGGAAATCTAGCTCAACGGA 48350

358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgG1 374
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

48351 ATPACACATCACGGGCAACAACAAACCCGACGAAGTCGTCGCCGCTGA 48400

374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
||||:||||:||||:||||: ||::: |||||:||||: ||::: |||||

48401 ATTACGCCAAATGGAATCCGCAACCTTACGACACCTTCCAAGCTGCAACGTT 48450

391 eArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

48451 CCAAAGACGGGTGCGAGTGTGGTGCTACTTCGACAAATGTCAGTTTCACT 48500

408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
::: ||||: :::::||||:||||:||||:||||:||||: ||::: |||||

48501 GGTGTCCCGCTTCCGGCACGCCGCAAAAGTCGATTTGAACATGAGTCT 48550

424 IolGluGlnProSerGlySerSerThrIleAlaIleGlyTyrSerGlnS 441
||||: ||||: :::::||||:||||: ||||| |||||:||||: |||||

48551 GACCCAACGTTCCACCGGTTCCCTGGATTGAGCGCGGGTGGGTTCAAG 48600

441 eRGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
:: ||||: :::::||||:||||: ||||| |||||:||||: |||||

48601 ATACCGGGTTGGTCATGTCGCGAGCGGTTCCCAAGACACACTGTCGGT 48650

458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
||||||| ||||| ::::: |||||:||||: |||||

48651 ACGGCAAGTCGGCGCACTGGCGCCTTCAGAGACAAACACCGCTTAA 48700

474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

48701 CGGCTCGCTGCTGTTACTGACCGCTACTTCAGCGCACACGGGTACGCC 48750

491 InSerLeuSerGlyTyrTyrArg..... LysThrLysTyrAspAsnLys 505
::: ::::: ||||| ||||| ::::: |||||

48751 TGGGCTACGATGTTTACGGAAAGCGCTTCGACCCGCGAAAGCATCGACC 48800

506 AsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrG1 522
:::||||:||||:||||: ||||| |||||:||||: |||||

48801 AGCATCAACAATATAAACCAACCGCAGCGCGCGAGCATCCCGCATGAG 48850

522 yTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
||||: ||||| ||||| ||||| ||||| ||||| ||||| |||||

48851 CGTGCTGTACCGAATACGACCGGCTGAATTCGGTTTGGTGGCAGAAC 48900

539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
:: ||||| ||||| ::::: |||||:||||: |||||

48901 ACCTGACCGCTCAAC.....ACCTCAACAACAAAGCGCCCAA 48935

556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAs 572
||||| ||||| ::::: |||||:||||: |||||

48936 CACTATGCGCACTTTATCAAGAATAACGGCAAAACCGACGGC...ACAGA 48982

572 pPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrPasnTyrS 589
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||

48983 CGGCAGCTTCAAGGCTGGCTGTACAAAGGTACCGTGGCTGGGGCCCA 49032

589 eRerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||

49033 ACAAAACCGCAGCGCGTTATGGCGACGCGCGGTACTCTCACGGCGGTG 49082

```
606 AspleuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValva 620
    ::::
    ::::
49083 AACGCCGAATCGCCTCGCTGGCAGCAACTGCAATACTACTCGCCAC 49132
    ::::
    ::::
620 ltyrGlnGlyAsnIleTyArgProPheIleLysSerValLeuArg. 636
    ::::
    ::::
49133 CCACAACCAACCTGGTTCCTCCCGCTGAGCAAAACCTTCACGCTGATGC 49182
    ::::
    ::::
637 .....GlyTyAlaLysLeuGlyTyArgLys.....AsnAsn 646
    ::::
    ::::
49183 TCGCGCGCAAGTCGGCATTCGGGC...GGCTACGGCAGAACCAAGAA 49229
    ::::
    ::::
647 LeuProPheTyGluAsnPheTyAlaGlyGlyTyArgLysSerValArgG1 663
    ::::
    ::::
49230 ATCCCTCTCTTGAATACTTCTACGGCGCGCTGGTTCGGTTCGGTGC 49279
    ::::
    ::::
663 yTyAspGlnSerSerLeuGlyProArgSerGlnAlaTyLeuThrAlaA 680
    ::::
    ::::
49280 ATCGAAGCGCGCAGCTCGCTCGGAAA..... 49307
    ::::
    ::::
680 rArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
    ::::
    ::::
49308 .....GTCTATCAGCAATACGGCGAATAATCAGTACGGCGGCAAC 49349
    ::::
    ::::
695 AlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheLysG1 711
    ::::
    ::::
49350 AAAAAAGCAACAGTCTCGCGCGAGCTGCTTCCCGATGCCGCGCGAA 49399
    ::::
    ::::
711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlnV 728
    ::::
    ::::
49400 AGAC....CGCGCGCAGTCCCGCTGAGCGCTGTTGGCAGCGACGACG 49446
    ::::
    ::::
728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
    ::::
    ::::
49447 TGTGGCGCGCAAAACCTACGACGACACACAGCTTCGCGCACC..... 49490
    ::::
    ::::
745 LysAspProGlnAlaThrAlaGluGlnAlaLysAlaAlaAsnArgPr 761
    ::::
    ::::
49491 .....GCGCGCAGGGTTCAAAACATTTACGCGCGCGCAATACCCA 49531
    ::::
    ::::
761 oLeuLeuThrGlnAspLysGlnLeuArgTySerAlaGlyValGlyAlat 778
    ::::
    ::::
49532 TAAATCCACCTTTACCAACGAATTCGCTATTCGCGCGCGCGGTTA 49581
    ::::
    ::::
778 hrTrpTyThrProIleGlyProLeuSerIleSerTyAlaLysProLeu 794
    ::::
    ::::
49582 CTGGCTCTCGCTTTAGCGCGGATGAAATTCAGCTACGCTACCGCGTG 49631
    ::::
    ::::
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
    ::::
    ::::
49632 AAGAAAAACCGGAAGACGAAATCCACGCTTCCCAATTCACCTCGGCAC 49681
    ::::
    ::::
811 rValPhe 813
    ::::
49682 GACGTTTC 49688
```

seq_name: /STDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAF21544

seq_documentation_block:

ID AAF21544 standard; DNA; 349980 BP.

XX AAF21544;

XX 13-MAR-2001 (first entry)

XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:1.

XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
ds.

XX Neisseria meningitidis.

OS

XX

PN WO200066791-A1.

XX 09-NOV-2000.

XX 08-MAR-2000; 2000WO-US05928.

XX 30-APR-1999; 99US-0132068.

XX 08-OCT-1999; 99WO-US23573.

XX 28-FEB-2000; 2000GB-0004695.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Pizsa M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;

PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;

PI Frazer CM, Grandi G;

XX WPI; 2000-647603/62.

XX Neisseria meningitidis B full length genome sequence and open reading

PT frames are used to detect, treat and prevent Neisserial infections -

XX Claim 7; Appendix A; 692pp; English.

XX The present invention describes the full length genome of

CC Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607

CC to AAF21613 represent fragments of the NMB genomic sequence, as the

CC sequence was too long to go in a record on its own it was split into 8

CC sequences which overlap each other at the beginning and end of each

CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at

CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at

CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the

CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to

CC AAF21606 represent PCR primers which are used in the exemplification of

CC the present invention. The NMB genome and fragments from it have

CC antibacterial activity, and can be used in vaccines and gene therapy.

CC Neisseria nucleic acids, proteins and/or antibodies which binds to the

CC proteins can be used in compositions for treating or preventing infection

CC due to Neisserial bacteria or as a diagnostic reagent for detecting the

CC presence of Neisserial bacteria or of antibodies raised to Neisserial

CC bacteria. Computers, computer memory, computer storage medium or computer

CC databases can be used in a search to identify open reading frames (ORFs)

CC or coding sequences within the NMB genome. The DNA sequences provide

CC further opportunities to find antigenic or immunogenic proteins which are

CC more effective in vaccines than the outer membrane proteins currently

XX used.

SQ Sequence 349980 BP; 83241 A; 85091 C; 95206 G; 86442 T; 0 other;

alignment_scores:

Quality: 1115.00

Ratio: 2.042

Percent Similarity: 66.667

Percent Identity: 32.479

Length: 819

Gaps: 14

alignment_block:

US-09-701-711-2 x AAF21544/rev ..

Align seg 1/1 to reverse of: AAF21544 from: 1 to: 349980

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26

180803 AAACCTGAACAGATTGCTTCGCACTGATGATGTTGGCATATCGCCTTT 180754

26 aGlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43

180753 GGCACCTGGCCGACATCCACATCCAGACATCCGCGTCGAGAGCTTGCGAGC 180704

43 rGValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59

180703 GTACCGAGCGGAGTACCGTATTCACTACCTGCGCCGCTCAAGTCGCGGAC 180654

728 alphaAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
 2150 TGTGGCGCGCAAAACCTACGACGACACAGCAGTTCGCGGACCC.....2193
 745 LysAspProGlnAlaThrAlaGlnAsnAlaLysAlaAlaAsnArgPr 761
 2194GGCGGCGGGTTCAAACATTTACGGCGCGCGCAATACCCA 2234
 761 OleuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
 2235 TAAATCCACCTTACCAACGAATTGCGCTATTCGCGCGCGCGGGTGA 2284
 778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
 2285 COTGGCTCTCGCTTTAGGCGCGGATGAATTCAGTACGCTACCGCGTG 2334
 795 AsnLysLysGlnAsnAspClnThrAspThrValGlnPheGlnIleGlySe 811
 2335 AAGAAAAACCGAAGAGAAATCCACGCTTCCAAATTCACACTCGGCAC 2384
 811 rValPhe 813
 2385 GACGTTTC 2391

seq_name: /SIBS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH42130

seq_documentation_block:

ID_AA42130 standard; DNA: 2394 BP.

AC_AA42130;

XX 17-SEP-2001 (first entry)

XX Nucleotide sequence of a Neisseria serogroup A protein.

XX Serogroup A protein; outer membrane protein; Neisserial infection;
 XX vaccine; ss.

XX Neisseria meningitidis.

XX Key Location/Qualifiers

FT CDS 1..2394

FT /*tag= a /product= "Neisseria serogroup A protein"

FT 1..63

FT /*tag= b

FT mat_peptide 64..2391

FT /*tag= c

XX WO200152885-A1.

XX 26-JUL-2001.

XX 17-JAN-2001; 2001WO-IB00166.

XX 17-JAN-2000; 2000GB-0001067.

XX 09-MAR-2000; 2000GB-0005699.

XX (CHIR-) CHIRON SPA.

XX Pizza M, Rappuoli R, Giuliani M;

XX WPI; 2001-451895/48.

XX P-PSDB; AAB84746.

XX Composition for treating or preventing infection to, detecting, or for
 PT raising antibodies against Neisserial bacteria, comprises an N.
 PT meningitidis serogroup B outer membrane preparation and an immunogenic
 PT component -
 XX Disclosure; Page 70-71; 83pp; English.

PS The present sequence encodes a Neisseria serogroup A protein. The

CC

CC protein is used to produce the compositions of the invention. The
 CC specification describes a composition, comprising a Neisseria
 CC meningitidis serogroup B outer membrane preparation and an immunogenic
 CC component. The immunogenic component is protein disclosed in WO99/57280,
 CC WO99/36544, WO99/24578, WO99/66791, WO97/28273, WO96/29412, WO95/03413,
 CC WO99/31132, WO99/58683, WO99/55873, and/or N. meningitidis protein PorA,
 CC TpaA, TbpB, PilC, Opa, or Omp85. The composition is used for making
 CC a medicament for treating or preventing infection due to Neisserial
 CC bacteria; a diagnostic reagent for detecting the presence of Neisserial
 CC bacteria; or of antibodies raised against Neisserial bacteria; and/or
 CC a reagent which can raise antibodies against Neisserial bacteria. It may
 CC also be used as a vaccine.

XX
 SQ Sequence 2394 BP; 655 A; 758 C; 586 G; 395 T; 0 other;

alignment_scores:

Quality: 1111.00 Length: 819

Ratio: 2.035 Gaps: 14

Percent Similarity: 66.667 Percent Identity: 32.357

alignment_block:

US-09-701-711-2 x AAH42130 ..

Align seg 1/1 to: AAH42130 from: 1 to: 2394

10 GlnValSerAlaMetThrMetAlaValMetMetValMetSerThrHisAl 26

4 AAACGTGAACACAGATTGCTTCGCCACTGATGTCTTGGGCATATCCGCTTT 53

26 acLnaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnA 43

54 GGCACCTTCGCACTTCCACATCCAAAGACATCCGCTCGAAGGCTTCGACG 103

43 rgValThrIleGluSerLeuGlnSerValLeuProPheArgLeuGlyGln 59

104 GTACGAGCCGAGTACCTGATTTCAACTACCTGCGCTCAAGTCGGCGAC 153

60 ValValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaTh 76

154 ACCTACACGACACACGCGCAGTGCCTCATCAAAAGCCTGTACGCCAC 203

76 rGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArgIleIle 93

204 CGGTTCCTTTGACGACGTACGCTCGCAAACTGCGGACGCGGACGCTCCTGC 253

93 yrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsn 109

254 TGACGCTTATCGACGCGCCACCACATCGGCTCGCTCAACATCACCGCGCA 303

110 ArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLe 126

304 AAAATGCTGCAAAACGACGCCATTAAGAAAACCTCGAATCGTTCGGCT 353

126 uAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrG 143

354 GCGCAGTCGCAATACCTTTAATCAGCGACACATCAATCAGGCGAGTCGCGC 403

143 luLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThr 159

404 GCCTGAAAGAGATACCTCGGCGCGGCAAACTCAATATCCAAATCAGC 453

160 ValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPh 176

454 CCCAAAGTAACCAAACTCGCGCGCAACCGCTCGACATCGACATCAGCAT 503

176 eAlaGluGlyLysProAlaArgValValAspIleAsnIleGlyAsnG 193

504 TGACGAGGCGCAATCCGCAAAATCAGCGACATCGAATTTGAAGCGCAACC 553

193 lnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsn 209

554 AAGTCTATTCGCGCAAACTGATGCGGCGAGATGCTGTCGACCGAAGGC 603


```

854 AAAAACTGCTGACCATGAAGCCCGGCAAAATGGTACGAACCCACAGCATG 903
308 GluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyrTy 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
904 ACCGCGTTTGGTGAGATTACAGACCGCATGGCTCGCAGCGTACGC 953
324 rTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrV 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
954 ATACAGCGAAATCAGGCTACAGCGCTGCCAAACCGCGCAACCAAAACCG 1003
341 alaValGluTyrTyrIleAspProValHisProValTyrValArgArg 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1004 TCATTTGCTGCTGCATCGAATCCGACCGCGGCGGAAATCTACGTCAAGAA 1053
358 IleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgG 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1054 ATCCACATCACCGCAACAACAAACCCGCGACGAGTGTGTCGCGCGCA 1103
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuS 391
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1104 ATTGCGCAAAATGGAATCCGCGCTTACGACACCTCCAAAGCTGCAAGCT 1153
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValasp 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1154 CCAAGAGCGCGTCCGAGCTTTTGGCTACTTCGACACGCTACAGTTTGAT 1203
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1204 CGCTCCCGCTTCCGCGCACACCGACAAAGTCGATTGACATGAGCCT 1253
424 lGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlnS 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1254 GACCGAACGTTCCACCGGCTCGCTGACTTGAGCGGGGTGGGTACAGG 1303
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1304 ATACCGGCTTGGTTCATGTCGCGAGGCGTTTCCCAAGACAACCTGTTCGGT 1353
458 ThrGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1354 ACGGCAAGTCGCGCGCTTCCGCGCTCACGAAGCAAAACACGCTCAA 1403
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1404 CGGCTCGCTCGCTTACACCGCGTACTTACCGCAGACGGGTGACGCC 1453
491 lnsrLeuSerGlyTyrTyrArg. ....LysThrLysTyrAspAsnLys 505
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1454 TGGGCTACGATGTTTACGGAAGCCCTTCGACCCGCGCAAGCATCGACC 1503
506 AsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrG 522
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1504 AGCATCAACAATATAAACCCACCGCAGCGGCGCAGGCATCCGATGAG 1553
522 yTyrProIleAspGlnAsnGlnArgIleSerPheGlyLeuAsnAlaAspA 539
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1554 CGTCCCTGTACCGAATACACCGCGTGAATTTTCGTTTGGTGGCAGAAC 1603
539 snThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGln 555
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1604 ACCTGACCGCTCAAC. ....ACCTACAACAAGCGCCCAAA 1638
556 LeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsnGlyIleProAs 572
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1639 CACTATGCCGACATTTATCAGAATAACGGCAAAACCGACGGC...ACAGA 1685
572 pPheLysHisAspTyrThrThrThrAsnAlaIleLeuGlyTyrTrpAsnTr 589
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1686 CGCAGCTTCAAAGGTGGCTGTACAAAGGTACCGTCCGCTGGGGCGCA 1735
589 erSerLeuAspArgProValPheProThrGlnGlyMetSerHisSerVal 605
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1736 ACAAACCCGACACGGGTTATGCCGACGCGGCTACCTGACGGGCGTG 1785

```

```

606 AspLeuThrValGlyPhe...GlyAspLys...ThrHisGlnLysValVa 620
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1786 AACGCCGAATCGCCCTGCCGCGCAGCAAACTGCAATACTACTCCGCCAC 1835
620 lTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLeuArg. 636
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1836 CCACCAACCAACCTGTCTTCCCTTAAGCAAAACCTTACGCTGATGC 1885
637 .....GlyTyrAlaLysLeuGlyTyrGly.....AsnAsn 646
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1886 TCGCGCGCAAGTCGCATTTGCGGC...GGCTACGGCAGAACCAAGAA 1932
647 LeuProPheTyrGluAsnPheTyrAlaGlyTyrGlySerValArgG 663
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1933 ATCCCTTCTTTGAAACTTCTACGGCGGCGCTGGTTCGGTGGCGG 1982
663 yTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaA 680
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1983 ATACGAAAGCGCAGCTCGTCCGAAA..... 2010
680 rArgGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsn 694
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2011 .....GTGTATGACGAATACGGCGCAAAATCAGCTACGGCGGCAAC 2052
695 AlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLysG 711
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2053 AAAAAAGCCAACTCTCCGCGAGCTGCTTCCCGATGCCTGGCGGAA 2102
711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlnV 728
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2103 AGAC...GCGCGCACCTCGCTGAGCCTGTTTGGCGGCGGCGGAGCG 2149
728 alPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2150 TGGGGACGGCAAAACCTACGACCAACAGCAGCTTCCGCGACC..... 2193
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2194 .....GGCGCAGGTTCAAAACATTTACGGCGCGCGCAATACCCA 2234
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2235 TAAATCCACCTTTACCAACGAATTCGCTATTTCGCGGCGGCGCGGTTA 2284
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2285 CTTGGCTCTCGCTTTAGGCGCGGATGAATTCAGGTACGCTACCCGCTG 2334
795 AsnLysLysGlnAsnAspGlnThrValGlnPheGlnIleGlyse 811
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2335 AAGAAAAACCGGAAGACAGAAATCCAACGCTTCCAATCCAACCTCGCAC 2384
811 rValPhe 813
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
2385 GAGCTTC 2391

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000. DAT:AAA15155

seq_documentation_block:

ID AAA15155 standard; DNA; 2505 BP.

XX AAA15155;

AC AAA15155;

XX 21-AUG-2000 (first entry)

DT 21-AUG-2000 (first entry)

XX DNA encoding outer membrane protein (omp) 85.

DE Outer membrane protein; omp85; gonococcal infection; symptomatic disease;

XX meningococcal infection; protective immune response; vaccine; ss.

KW Neisseria gonorrhoea.

XX

OS

XX


```

1113 ATCCACATCCGCGCAACAACAAACCGGACGAAGTCGTGCGCGCGA 1162
374 uMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLys 391
1163 ATTGCGCAATGGATCCGCGCTTACGACACTCTCAAGCTGCACGCT 1212
391 erArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAsp 407
1213 CCAAGAGCGCTCGAGCTTTGGGTACTTCGACAACTGACAGTACAGT 1262
408 ThrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 424
1263 GCGTCCCGCTTCGCGGTACGCGCGCAAAAGTCGATTGAACATGAGCCT 1312
424 lGluGluGlnProSerGlySerThrIleAlaAlaGlyTyrSerGlns 441
1313 GACCGAACGTTCCACGCGCTCGCTGACCTTGACGGCGGTGGTTCAGG 1362
441 erGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPheMetGly 457
1363 ATACCGCTTGGTCACTCCGCGCGGTATFCGACGAGCAACCTGTTCCGT 1412
458 ThrGlyIysHisValAsnAlaSerPheSerArgSerGluThrArgGluVa 474
1413 ACGGCAAGTCGCGCGCTCGCGCGCTCGCGAAGCAAAACCAACGCTCAA 1462
474 lTyrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerG 491
1463 CGCTCGCTGCTGTTACCGACCGGTACTTCACGCGACGCGGGTCAGC. 1511
491 lnSerLeuSerGlyTyr.....TyrArgLysThr.....LysTyr 502
1512 .....CTGGCTACGATATTACGGAAAGCCTTCGACCGCGCGCAA 1553
503 AspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlySerLe 519
1554 GCATGACACGCTCAACAAATAAATAAACACACCGCGCGCGCGGT 1603
519 uSerTyrGlyTyrProIleAspGlnGlnArgIleSerPheGlyLeuA 536
1604 AAGGATGGGTATCCCGTTACCGAATACGACCGCGTCAATTCGGGTGG 1653
536 snAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsn 552
1654 CGCGGAACACCTGACCGTCAAC.....ACCTACAACAAA 1688
553 valLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnGln 569
1689 GCACCAAAAGCTATCCGACTTTATCAACAATACGGCAAAACCGACGG 1738
569 ylleProAspPheLysHisAspTyrThrTyrAsnAlaIleLeuGlyT 586
1739 C....GCAGACGGCAGCTTCAAGCGCTGTGTACAAAGGCACTGTGCGCT 1785
586 rPAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSer 602
1786 GGGGCGCAACAAGACCGACGCGCTTATGGCGCGACGCGCGCTACCTG 1835
603 HisSerValAspLeuThrValGlyPhe...GlyAspLys...ThrHisG 617
1836 ACCGGCGTAATCCGAATCGCCCTGCGCGGACGAAACTGCAATACTA 1885
617 nLysValValTyrGlnGlyIleTyrArgProPheIleLysLysSerV 634
1886 CTCGGCCACCAACAACCTGGTCTTCCCTTAAGCAAAACCTTCA 1935
634 alLeuArg.....GlyTyrAlaLysLeuGlyTyrGly... 644
1936 CGTGATGCTCGCGCGGAAGTCGGCATTCGGGC...GGCTACGGCAGA 1982
645 ...AsnAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlyse 660

```

```

1983 ACCAAGAAATCCCTTCTTTGAAAACTTCTACGGCGCGGCTGGGTTCC 2032
660 rValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrL 677
2033 GGTGCGGGCTACGAAGCGGACGCTCGGCCGAAA..... 2069
677 euThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluValVal..... 691
2070 .....GTGTATGACGAATACGGCAAAATCAGCTAC 2102
692 GlyGlyAsnAlaLeuAlaThrPheGlySerGlyLeuIleLeuProLeuPr 708
2103 GGGGGAACAAACAAAGCAACGCTCGCGGAGCTCTTCCCGATGCC 2152
708 oPheLysGlyAspTyrIleAspGlnValArgProValIlePheIleGluG 725
2153 CGGTGCGAAAGAC...GCAGCACCGCTCGCGCTGAGCTGTTTCCCGACG 2199
725 lyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeu 741
2200 CAGCGACGCTGTGG.....GACGGCAGAACCTATACCGCC 2234
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaA 758
2235 GCCGAAACGTAACAAATCGTTTACTCGGAAACGCGATAAATC 2284
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
2285 C.....ACCTTACCAACGAATGCGCTATTTCCCGCGCG 2319
775 alGlyAlaThrTyrThrProIleGlyProLeuSerIleSerTyrAla 791
2320 CGCGGTTACCTGCTCTCGCTTTGGCGCGGATGAAATTCATCTACGCC 2369
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG 808
2370 TACCGCTGAAGAAAAACCGAAGACGAAATCCACGCTTCCAATTTCCA 2419
808 nIleGlySerValPhe 813
2420 GCTCGGACGACGTTCC 2435

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-n-emb1/NA1994.DAT:AAQ66199

seq_documentation_block:

ID AAQ66199 standard; DNA; 2984 BP.

XX AC AAQ66199;

XX DT 08-DEC-1994 (first entry)

XX DE H. influenzae b Eagen D15 sequence.

XX KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;

XX KW OMP; outer membrane protein; Hib; ss.

XX OS Haemophilus influenzae type b Eagen strain.

XX FH Key Location/Qualifiers

XX FT CDS 374..2767

XX XX /*tag= a

XX PN W09412641-A.

XX PD 09-JUN-1994.

XX PF 23-NOV-1993; 93WO-CA00501.

XX PR 23-NOV-1992; 92GB-0024584.

XX PA (CONN-) CONNAUGHT LAB LTD.

XX PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;

PI Yang Y;
 XX WPI; 1994-200269/24.
 DR P-PSDB; AAR53755.
 XX
 PT Nucleic acid encoding D15 outer membrane protein - esp. of
 PT Haemophilus influenzae, and related proteins, vectors, antisera
 PT etc. useful in vaccines, for diagnosis and for passive
 PT immunisation.
 XX
 PS Disclosure; Fig. 1B; 16lpp; English.
 XX
 CC Outer membrane protein (OMP) D15 genes were isolated by screening
 CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
 CC Egan and Mann A, and the non-typeable (NTHi) strains SB33 and PAK
 CC 12085. Nucleotide sequences were determined for the D15 genes
 CC (AAQ66198-202) and the corresponding aa sequences were derived
 CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
 CC free of other antigens and lipooligosaccharides, by recombinant DNA
 CC methods using the isolated genes.
 XX
 SQ Sequence 2984 BP; 960 A; 474 C; 619 G; 931 T; 0 other;

alignment_scores:

Quality: 1052.50 Length: 822
 Ratio: 1.956 Gaps: 20
 Percent Similarity: 65.450 Percent Identity: 31.630

alignment_block:

US-09-701-711-2 x AAQ66199 ..

Align seg 1/1 to: AAQ66199 from: 1 to: 2984

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleth 39
 413 ACAACGACTGCTGTTCGCCGACTTGTGGCAAGAGATATTTCGTGGA 462
 39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
 463 TGGTGTCAAGTGACTTAGAACAACAATCCGAGCAAGTTTACCTGTC 512
 56 rGlyGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVallysAla 72
 513 GTCCCGGTGACGTGTGATGATGATGATGATGATGATGATGATGATG 562
 73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
 563 TTATTCGTAAGTGGTTCGATTCGATGATGATGATGATGATGATGAT 609
 89 yargile...ileTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
 610 CGATGCTGCTGTTGTTAGGCTTGGCTAAATCGATCAATTCAGATGTA 659
 105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
 660 AAATCAAGGTAACTCTGTATTTCCTGCTGATGATGATGATGATGAT 709
 122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrVal 138
 710 GATGCTACCGGTTTAAAGTTCGATGATGATGATGATGATGATGATG 759
 138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrA 155
 760 TGAATTTGCCAAAGTAAAGAGACACTATCGAAGTGTAGTTCGCTATA 809
 155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgVallys 171
 810 ACGCAACAGTGTGAACCTATTGTCAATACGCTACCAATAAATACGCT 859
 172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIle 188
 860 ATTTTAAATCAATCAATGAAGATGATGAAGCAAAATTCGATCATTA 909

188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
 910 TTTCAAGGGGAACGAATCTGTTAGTAGCAGTACATTACAAGAACAAATG 959
 205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
 960 AATTACAACCTGATCTCTGGTGGAAATTA...TGGGAAATAAATTTGAA 1006
 222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
 1007 GTGCGCAATTCGAAAGATTTCAGTCAATTCGATGATGATGATGATGAT 1056
 238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
 1057 TAATGGCTATGCCAAGACCAAAATTAATAAACGGATGTTACAGTAAATG 1106
 255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGln 271
 1107 ATGAAAAACAAGATTAAATGAACATTGATGATGATGATGATGATGAT 1156
 272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
 1157 TATGACCTTCGTAGTCAGCAATTAATGATGATGATGATGATGATGATG 1206
 288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
 1207 CGAGCTTGAACCTTTACTTTTACAGCATTAATTAATGATGATGATGAT 1256
 304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
 1257 GTAGTGATATTCAGATGTAGAAATTAATTAAGCAAACTTGGAGAA 1306
 321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG 337
 1307 CCGGTTCGTTAGCGCAACGTAATTCAGTACCTGATGATGATGATGATG 1356
 337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProVal 354
 1357 AAATAAAACATTAGCGATAACCTTGTGTTGATGCTGGACGACGTTTAA 1406
 354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
 1407 CTTGTCGCAACTTCGTTTGAAGAAATACCGTTTCTGCTGATGATGAT 1456
 371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGln 387
 1457 TTACGTCAGGAATGCGCCCAACAGAGGAAGTGGTATTAATTCACAAT 1506
 387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHis 404
 1507 ACTTGAGTTAGGAAAAATTCGCTTAGCTAGCTAGCTAGCTAGCTAGCT 1552
 404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
 1553 ..ACAGTCGAAACCGAATGATGATGATGATGATGATGATGATGATGAT 1600
 419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
 1601 GATGTCGATATTAAGTCAAGACGTAACACGGGTAGTATCAACTTGG 1650
 435 aAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspValSerG 452
 1651 TATTGTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTGTAAC 1700
 452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
 1701 AAGATAATTTCTGGCAACAGGGGGCGGCAAGTATAGTATGATGATGAT 1750
 469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
 1751 AATGATTATGTTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTAC 1800

```
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
1801 TAAAGATGGTGA.....AGTCTTGGGAAATGTTTCTTTGAAAACCT 1844
502 yAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
1845 AGGATAACTCTAAAGTGATACATCTCTPACTATPAAAGCGTAGCATTAC 1894
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
1895 GGAAGTAATGTTACTTGTAGGTTCCCTGTAATGAAATAAATCCCTATTA 1944
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
1945 TGTAGGATTAGGTACATACCTATAATAAATAGTAACCTTGTCTAGAAT 1994
544 lYcLYArgPheMetGlyIleSerAsnValLysGlnLeuMetaLAspGly 560
1995 ATAACCGTAATTATATATATCAATCAATGAA..... 2026
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
2027 .....TTTAAAGGTAATGCCATTAACAAACAAATGACTTTGTATT 2064
577 rThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspArgP 594
2065 TTCT.....TTTGGTTGGAACTATACAGCCCTTAATAGAG 2099
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
2100 GCTATTCCCACTAAAGGGGTTAAAGCAAGCTTGGTGGACGAGTTACT 2149
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
2150 ATTCCAGGTTCTGATAACAATACTACAACTAAAGTGCAGATGTACAGG 2199
625 eTyArgProPhe.....IleLysLysSerValL 635
2200 TTCTACCCCTTAGACAGAGACACCTCTGGGTTGTATCTGCAAAAGCAT 2249
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn....AsnLeuProPheTyr 650
2250 CTGCAGGATGCAAA...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2296
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
2297 CAAACTTATACAGCGGGTGGCATCGGTTACCTACGTTGTTTGTATGG 2346
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
2347 TAGTATTGACCTAAGCA.....ATTATGCCGAATATGGTAATG 2387
684 lNThrThrLeuGly.....GluValValGlyGly 693
2388 GTAGTGGTACTGTACTTTTAAAGAAGATAAGTCTGATGTGATTGGTGT 2437
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
2438 AATGCAATCGCTACAGCTAGCCAGAGTAAATGTGCCAACCTCCATTTCT 2487
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
2488 GAGCGATAGACCCAAATACGCTCGGAACCTCTTATTTGTGTATGCGG 2537
726 lYcLInValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
2538 CAAGTGTTTGGAATACTAAATGAAATCAGATAAAATGGATTAGAGC 2587
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
2588 GATGATTAAAA..... 2599
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
```

```
... |||
2600 .AGATTGCCTGATTATGGCAAAATCAAGCCGATTATCGCGCTCTACAGGTG 2648
775 aGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
2649 TCGGATTCCAAATGGCAATCTCTATTTGGGCCATTGGTATTTCTTATGCC 2698
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG 808
2699 AAACCAATTAATAAATATGAAATATGATGATGTCGACAGTTCCAATTTAG 2748
808 nIleGlySerValPhe 813
2749 TATTGGAGGTTCTTTC 2764
seq_name: /SID51/gcdata/geneseq/geneseq-emb1/NA1994.DAT:AAQ66198
seq_documentation_block:
ID AAQ66198 standard; DNA; 2949 BP.
XX
AC AAQ66198;
XX
DT 08-DEC-1994 (first entry)
XX
DE H. influenzae b Ca D15 sequence.
XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; Hib; ss.
XX
OS Haemophilus influenzae type b Ca strain.
XX
FH Key Location/Qualifiers
FT CDS 75..2468
FT FT /*tag= a
PN W09412641-A.
XX
PD 09-JUN-1994.
XX
PF 23-NOV-1993; 93WO-CA00501.
XX
PR 23-NOV-1992; 93GB-0024584.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
XX
DR WPI; 1994-200269/24.
DR P-PSDB; AAR53754.
XX
PT Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
PT etc. useful in vaccines, for diagnosis and for passive
immunisation.
XX
PS Disclosure; Fig. 1A; 161pp; English.
XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typeable (NTHi) strains SB33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
SQ Sequence 2949 BP; 983 A; 477 C; 612 G; 877 T; 0 other;
```

```
alignment_scores:
Quality: 1051.50      Length: 822
Ratio: 1.958          Gaps: 20
```

[illegible]

908 CGAGCTTGAACCTTTACTTTCAGCATTTACATNTTAATGATACTTTCGCC 957

304 InalameLeuGlInThrThrAsnAsnIleSerThrLysPheGlyasp 320
::: ::::: ||| ||||:::||::||::||::||::||::||::||
958 GTAGTGATATTGCAGATTAGAAAATGCAATTAAGAACAATACTGGAGAA 1007

321 AspGlyTyrtYrAlaGlnIleArgProValthrargIleAsnAspGI 337
||||| ||| ::::: ||| ::::: ||| ::::: |||
1008 CCGGGTTACGGTAGCCAACGGTAATTCAGTACTCTGATTTTGATGATGC 1057

337 uSerArgThrValAspValGluTyrtYrIleAspProValHisProValt 354
:::::||||::: ||| ::::: ||| ::::: ||| ::::: |||
1058 AAATAAACAATTAGCGATAACCCTTGTTGTATGTCGGCAGCAGCTTAA 1107

354 yrValargArgIleAsnPheThrGlyAsnPheLysThrInAspGluVal 370
|||||:::||:::||:::||:::||:::||:::||:::||:::||:::||
1108 CTGTTTCGCAACTTCGCTTGGAAGAAAAATCCGTTTCTGCTAGTACACT 1157

371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
|||||:::||:::||:::||:::||:::||:::||:::||:::||:::||
1158 TTACGTTCAGAAATCGCCACAGAAGAACTTGGTATTAATTCACAATT 1207

387 stIeGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHiv 404
::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::
1208 AGTTGAGTTAGGAAAAATTCGCTAGATCGTCAGGTTTCTTCGAA.... 1253

404 alThrValaspThrArg.....ProValProAsnSerProAspGlnVal 418
1254 ..ACAGTCGAAAACCGAATTGATCCTATCAATGGTAGTATGATGAAGAT 1301

419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
||||| ::::: ||||:::||:::||:::||:::||:::||:::||:::
1302 GATGTCGTATATAAGTCAAAGACGTAAACGGGTAGTATCAACTTTGG 1351

435 alaGlyTyrserGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
::: ||||::: |:::||:::||:::||:::||:::||:::||:::||:::
1352 TATTGGTTACGGTACAGAGAGTGGTATTAGTTATCAAGAAGTGTATAAC 1401

452 InaAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
|||||:::||:::||:::||:::||:::||:::||:::||:::||:::||
1402 AAGATAAITTTCTGGGAACAGGGCGGCAAGTAAGTATAGCTGGTACGAAA 1451

469 SerGluThrArgGluValTyrtSerLeuGlyMethThrAsnProtyrPheTh 485
::: ||||::: |:::||:::||:::||:::||:::||:::||:::||:::
1452 AATTGATTATGTCAGAGTGCAATTTGGGTTATACGAGCCCTATTTTAC 1501

485 rValasnglValserGlnSerLeuSerGlyTyrtYrArgLysThrLyst 502
| ::||::||| |:::||:::||:::||:::||:::||:::||:::||:::
1502 TAAAGATGGTGTA.....AGTCVTGGTGGAAATGTTTTCTTTGAAAACT 1545

502 tyrAspAsnLys.....AsnIleSerAntyrrValLeuAspSerTyr 515
::: ||||::: |:::||:::||:::||:::||:::||:::||:::||:::
1546 ACGATACTCAAAGTGATACATCCTAACTATATAAGCGTACGACTTAC 1595

516 GlyGlySerLeuSerTyrglyTytrProIleAspGluAsnGlnArgIleSe 532
::: ||||::: |:::||:::||:::||:::||:::||:::||:::||:::
1596 GGAAGTAATGTACTTTTAGGTTTCCCTGTAAATGAAATAACTCCATTAT 1645

532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
||||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1646 TGTAGCATTAGGTACATACCTATAATAAATTAGTPACTTTTGCTCTAGAAT 1695

544 lyGlyArgPheMetGlyIleSerAsnVallLysGlnLeuMetalAlaAspGly 560
::: ||||::: |:::||:::||:::||:::||:::||:::||:::||:::
1696 ATAACCGTAATTATATATTCAATCAATGAAA..... 1727

561 GlyLysIleGlnValAspAsnGlnIleProAspPheLysHisAspyr 577
::: ||||::: |:::||:::||:::||:::||:::||:::||:::||:::
1728TTTAAAGGTAATGGCATTTAAACAATAACACTTGTGATT 1765

577 rThrThrTyrAsnAlalleLeuGlytyrpsAntyrrSerSerLeuAspargp 594
::: ||||::: |:::||:::||:::||:::||:::||:::||:::||:::
1766 TTCT.....TTTGGTTCGGAAGTATAACAGGCTTAATAGAG 1800

```
594 roValPheProThrGlnClyMetSerHisSerValAsp.....LeuThr 608
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1801 GCATTTCCCACTAAAGGGTTAAAGCAAGTCTTGGTGACAGTTACT 1850
609 ValGlyPheGlyAspLysThrHisGlnLysValValThrGlnGlyAsnI 625
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1851 ATTCCAGGTTCTGATACAAATACTACAAATAAGTGCAGATGACAGG 1900
625 eTyrArgProPhe.....IleLysLysSerValL 635
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1901 TTTCTACCAATTAGACAGAGATCACCTCTGGGTGTATGCAAAAGCAT 1950
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1951 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTCTAT 1997
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1998 CAACTTATACAGCGGTGGCATCGGTTTACCTGTTTCTCTATGG 2047
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2048 TAGATGGACCTAACGCA.....AFTTATGCGGAATATGGTAATG 2088
684 InThrThrLeuGly.....GluValValGlyGly 693
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2089 GTAGTGTACTGTCTTTTAAAGACAGATAAGTCTGTATGATGGTGGT 2138
694 AsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPhe 710
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2139 AATGCAATCGCTACAGTAGCCAGAGTTAATGTGCCAACTCCATTTGT 2188
710 sGlyAspTTPle...AspGlnValArgProValIlePheIleGluGly 726
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2189 GACCGTAGAGCAAAATACGTGCGACCTCTTATTGTTGATGCGG 2238
726 lyGlnValPheAspThrT...GlyMetAspLysGlnThrIleAspLeu 741
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2239 CAAGTGTGGAATACTAATAATGAAATACAGATAAAATGATGAGAGC 2288
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2289 GATGATATAAA..... 2300
758 aAsnArgProLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2301 .AGATTGCTGATATTATGGCAATCAAGCGTATTCGGCCTCACAGTG 2349
775 alGlyAlaThrTTPThrProIleGlyProLeuSerIleSerTyrAla 791
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2350 TCGGATTCCAATGCCAATCTCTATTGGGCCATTGTTCTCTTATGCC 2399
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG 808
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2400 AAACCAATTAATAATGATAATGATGATGCGAAGTTCATTTAG 2449
808 nileGlySerValPhe 813
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2450 TATGAGGTCTTTC 2465
seq_name: /SID1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:AAQ66200
seq_documentation_block:
ID AAQ66200 standard; DNA; 2953 BP.
XX
AC AAQ66200;
XX
DT 08-DEC-1994 (first entry)
XX
DE H. influenzae b Minn A D15 sequence.
XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
```

```
KW OMP; outer membrane protein; Hib; ss.
XX
OS Haemophilus influenzae type b Minn A strain.
XX
FH Key Location/Qualifiers
FT CDS 334..2727
FT /tag= a
XX
XX WO9412641-A.
XX
XX 09-JUN-1994.
XX
XX 23-NOV-1993; 93WO-CA00501.
XX
XX 23-NOV-1992; 92GB-0024584.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
XX Yang Y;
XX
XX WPI: 1994-200269/24.
XX P-PSDB; AAR53756.
XX
XX Nucleic acid encoding D15 outer membrane protein - esp. of
XX Haemophilus influenzae, and related proteins, vectors, antisera
XX etc. useful in vaccines, for diagnosis and for passive
XX immunisation.
XX
XX Disclosure; Fig. 1C; 161pp; English.
XX
XX Outer membrane protein (OMP) D15 genes were isolated by screening
XX chromosomal libraries of H. influenzae type b (Hib) strains Ca, PAK
XX Eagan and Minn A, and the non-typeable (NTHI) strains SB33 and PAK
XX 12085. Nucleotide sequences were determined for the D15 genes
XX (AAQ66198-202) and the corresponding aa sequences were derived
XX (AAR53754- 58). D15 OMP can be produced easily and on a large scale,
XX free of other antigens and lipooligosaccharides, by recombinant DNA
XX methods using the isolated genes.
XX The three 'n's are not identified in the specification.
XX
XX Sequence 2953 BP; 948 A; 467 C; 619 G; 916 T; 3 other;

alignment_scores:
Quality: 1050.50 Length: 822
Ratio: 1.953 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x AAQ66200 ..
Align seg 1/1 to: AAQ66200 from: 1 to: 2953
23 SerThrHisAlaGlnAlaAspPheMetAlaAsnAspIleThrIleTh 39
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 ACAACGACTGTGTTGCGCACCTTTTGTGGCAAAAGATATTCGTGGA 422
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
423 TGGTGTTCAGAGTGACTTAGAACAAACAATCCGCAAGTTTACCTGTC 472
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
473 GTCCCGGTGACGCTGACGTGACATGATGCTGATATATTGTCCTCT 522
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
523 TTATTCGTAAAGTGTGATTCGATGATGTAACCG...CATCAAGAGG 569
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluLeA 105
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
570 CGATGTGCTGTTGTTAGCGTTGTGGTAAATCGATCATTCATTCAGATCTTA 619
```



```
...
2348 GTAGTGGTACTGGTACTTTTAAGAAGATAAGTCTCTGATGTGATGGTGGT 2397
694 AsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuPheLeu 710
2398 AATCAATCGCTACAGTAGCGCAGAGTTAATGTGCACTCCATTTGT 2447
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGly 726
2448 GAGCGATAAGAGCCAAATACGGTCCGAACTCTTATTGTTGATCGG 2497
726 lyGlnValPheAspThr...GlyMetAspLysGlnThrIleAspLeu 741
2498 CAAGTGTGGTGAATACATAAATGGAATACAGATAAAATGGATTAGAGC 2547
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758
2548 GATGTATTAAAA..... 2559
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGly 775
2560 .AGATTGCCTGATTATGCAAAATCAAGCGTATTTCGCGCCTCTACAGGT 2608
775 alGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
2609 TCGGATTCCCAATGGCAATCTCTATTGGCCATTGGTATTCTCTATGCC 2658
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG1 808
2659 AAACCAATTAAATAATGAAATGATGATGTCGAACAGTTCCAATTAG 2708
808 nIleGlySerValPhe 813
2709 TATTGGAGGTTCTTTC 2724
```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:AAQ66202

seq_documentation_block:

```
ID AAQ66202 standard; DNA; 2989 BP.
AC AAQ66202;
XX
DT 08-DEC-1994 (first entry)
XX
DE H. influenzae PAK 12085 D15 sequence.
XX
KW Vaccine; passive immunization; vector; antiserum; diagnosis; D15;
KW OMP; outer membrane protein; NTHi; ss.
XX
OS Haemophilus influenzae PAK 12085.
XX
FH Key
FT CDS
FT 389..2771
FT /*tag= a
XX
PN W09412641-A.
XX
PD 09-JUN-1994.
XX
PF 23-NOV-1993; 93WO-CA00501.
XX
PR 23-NOV-1992; 92GB-0024584.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Chong P, Klein M, Loosmore S, Sia DYC, Thomas W;
PI Yang Y;
XX
DR WPI: 1994-200269/24.
DR P-PSDB; AAR53758.
XX
PT Nucleic acid encoding D15 outer membrane protein - esp. of
PT Haemophilus influenzae, and related proteins, vectors, antisera
```

```
PT etc. useful in vaccines, for diagnosis and for passive
PT immunisation.
XX
PS Disclosure; Fig. 1E; 161pp; English.
XX
CC Outer membrane protein (OMP) D15 genes were isolated by screening
CC chromosomal libraries of H. influenzae type b (Hib) strains Ca,
CC Eagan and Minn A, and the non-typeable (NTHi) strains SR33 and PAK
CC 12085. Nucleotide sequences were determined for the D15 genes
CC (AAQ66198-202) and the corresponding aa sequences were derived
CC (AAR53754-58). D15 OMP can be produced easily and on a large scale,
CC free of other antigens and lipooligosaccharides, by recombinant DNA
CC methods using the isolated genes.
XX
SQ Sequence 2989 BP; 975 A; 468 C; 615 G; 931 T; 0 other;
```

alignment_scores:

```
Quality: 1047.50      Length: 818
Ratio: 1.943          Gaps: 21
Percent Similarity: 65.892 Percent Identity: 31.663
```

alignment_block:

US-09-701-711-2 x AAQ66202 ..

Align seg 1/1 to: AAQ66202 from: 1 to: 2989

```
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
429 ACAACGACTGTGTTGGCGCACCTTTTGTGCAAAAGATATTCGCTGGA 478
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
479 TGGTGTTCAGTGCTAGTAGAACACAAATCCGAGCAAGTTTACCTGTC 528
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
529 GTGCTGGTGGTGTGACTGACAATGATGTGGCTAATATTGTCGCTCT 578
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG1 89
579 TTATTCGTAAGTGTGCTGATGATGTGAAGCG...CATCAAGAAGG 625
89 YArgile...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
626 CGATGCTGCTGTTAGCTGTGGCTAATCGATCATTCAGATGTTA 675
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
676 AAATCAAGGTAACCTCTGTTATCCCACTGAAGCACTTAAACAAACTTA 725
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG1 138
726 GATGCTAACCGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTAAA 775
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTrA 155
776 TGAATTTGCCAAAAGTGTAAAAGAGCACTATGCAAGTGTAGTGCCTATA 825
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
826 ACGCAACCGTTGAACCTATTGTCAATACGCTGCCAATAATACGTGCTGAA 875
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
876 ATTTTAATTCAAATCAATGAAGATGATAAAGCAAAATTTGGCATTAAC 925
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
926 TTTCAGGGGAACGAATCTGTTAGTAGCATACATTACAGAAACAATGG 975
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
```

976 AATTACAACTCGATTCTTGGTGGAAATTA...TGGGAAATAAATTTGAA 1022
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
1023 GUTGGCAATTCGAGAAAGATCTCGAGCAATTCGTGATTATTATTAA 1072
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
1073 TAATGGCTATGCCAAGACAAATCACTAAACGGATGTTACGTAATATG 1122
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGln 271
1123 ATGAAAAACAAGATTAATGTAACCATTTAGTAAATGAAGTTTACAG 1172
272 TyrArgPheGlyGlnThrGlnPheLeuAsnLeuThrTyrThrGlnAl 288
1173 YATGACCTTCGTAGTCAGGCATTTATAGTAATCTGGGAGGTATGCTGC 1222
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
1223 CGAGCTTGAACCTTTACTTTACGATTTACATTTAAATGATACTTTCCGCC 1272
304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
1273 GTAGTGATATTCGATGATGAGAAATGCAATTAAGCAAAACTTGGGAA 1322
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG 337
1323 CGAGCTTACGGTAAACACACACAGTAAATCTGTACCTGATTTTGACGATGC 1372
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
1373 AAATAAACAATTAGGATACACCTTTGTTGTTGATGCTGGACGACGTTTAA 1422
354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
1423 CTGTTCCGCACTTCGCTTTGAAGGAATACCGTTTCGCTGATAGTACT 1472
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnly 387
1473 TTACGTCAGGAAATGCGACAACAAGAACTTGGTATATATCAAAAT 1522
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
1523 AGTTGAGTTAGGAAATAATTCGTTAGATCGTACAGGTTTCTTCGAA... 1568
404 alThrValAspThrArg...ProValProAsnSerProAspGlnVal 418
1569 ..ACAGTTGAAACCGAATTTGATCTATCAATGTTAGCAATGATGAAGTG 1616
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
1617 GATGTCGTATATAAAGTCAAGAAGCTAAACGGGTAGTATCAACTTTGG 1666
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
1667 TATGTTGACGGTACAGAGTGGTATCATGTTATCAACAAGTATTAAC 1716
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
1717 AGATAAATTTCTTGGACACGGGGCGGAGTAAGTATATAGCTGGTACGAAA 1766
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
1767 AATGATTATGTTACGAGTGTCAATTTGGGTTATACCGAACCCCTATTTTAC 1816
485 rValAsnGlyValSerGlnSerGlyTyrTyrArgLysThrLysT 502
1817 TAAAGATGGTGTA....AGTCTTGGTGGAAATATTTCTTTTGAAGAACT 1860
502 yzAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
1861 ACGATAACTCTAAAGTGATACATCTCTAACTATAAGCGTAGGACTTAT 1910

516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
1911 GGAAGTAAGTTACTTACCTTAGCTTCCCTGTAATGAAATAACTCTCTATTA 1960
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
1961 TGTAGGATTAGGCCATACCTATAATAAATAATAGTAACTTTGCTCTAGAAT 2010
544 lyGlyValArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
2011 ATAACCGTAATTTATATATTAATCAATCAATGAAA..... 2042
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
2043TTTAAAGGTAATGCAATTAACCAAAATGACTTTTGATTT 2080
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
2081 TTCT.....TTTGGTTGGAACATATAACAGCCTTAATAGAG 2115
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
2116 GCTATTTCCCACTAAAGGGTTAAAGCAAGTCTTGGTGGACGAGTTACT 2165
609 ValGlyPheGlyAspLysThrHisGlnLysValVal.....TyrGlnGl 623
2166 ATTCCAGGTTCTGTATAACAAATACTACAAACTAAAGTGCAGATGTACAGGG 2215
623 y.....AsnIleTyrArgProPheIleLysLysSerValL 635
2216 TTCTACCATTTAGACAGATACCGCTGGGTTGTATCTGCAAAAGCAT 2265
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
2266 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCCTTCTAT 2312
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
2313 CAACTTATACAGCGGTGTCATTTGTTTACGCGGTTTTCCTTATGG 2362
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
2363 TAGTATTTGGCCTTAATGCA.....ATTATGCCGAACATGGTAATG 2403
684 InThr.....ThrLeuGlyGluValValGlyGlyAsnAlaLeuAla 697
2404 GTACTTTTAAATAAGATAAGTCTCTGATGTGATGTTGTTAATGCAATCACA 2453
698 ThrPheGlySerGluLeuLeuLeuProLeuProPheLysGlyAspTrpIl 714
2454 ACTGCGAGTGCAGAACTTATGTACCAACTCCATTTGTGAGTGATAAAG 2503
714 e...AspGlnValArgProValIlePheIleGluGlyGlyGlnValPheA 730
2504 CCAAAATACAGCCCGAACCTCCCTATTGTTGATGCGGCAAGTGTGGGA 2553
730 spThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPheLys 745
2554 ATACTAAATGGAAATCAGATAAAATGGATTAGAGAGCAAGGCTTTGAAA 2603
746 AspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLe 762
2604 GAC.....TTACCTGA 2614
762 uLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrT 779
2615 TTATGGCAATCAAGCCGATTTCGCCCTCTACAGGTTGCGGATTCAT 2664
779 rPThrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsn 795
2665 GGCAATCTCTTATGGACCATTTGTTATTTCTTATGCTAAACCAATAAA 2714


```

1319 CGAGGTTACGGTAACACACAGATAAATCTGTACCTGATTTTGACGATGC 1368
      |||||      :      :      :      :      :      :      :      :      :
337 uSerArgThrValAspValGluTyrThrIleAspProValHisProValT 354
      :|||:      :      :      :      :      :      :      :      :
1369 AAATAAAACATTAGCAGTAACCTTTGTTGTGTGCTGGCAGCAGTTTAA 1418
      :|||:      :      :      :      :      :      :      :      :
354 yrvAlargArgIleAsnPhetheGlyAsnPhelystThrGlnAspGluVal 370
      |||||      :      :      :      :      :      :      :      :
1419 CTGTTCCACCAACTCGCTTGAAGGAATACCGTTTCTGCTAGTAGTACT 1468
      :|||:      :      :      :      :      :      :      :      :
371 LeuArgArgGluMetArgGlnLeuGluGluAlaLeuAlaSerAsnGlnLy 387
      |||||      :      :      :      :      :      :      :      :
1469 TTACGTCAGGAATGCGCCCAAGAAGAACTTGGTATAATTCACAATT 1518
      :|||:      :      :      :      :      :      :      :      :
387 sIleGlnLeuSerArgAlaLeuMetArgThrGlyPhePheLysHisV 404
      :|||:      :      :      :      :      :      :      :      :
1519 ACTTGAGTTAGGAAATAATCGCTAGATCGTACAGGTTTCTTCGAA... 1564
      :|||:      :      :      :      :      :      :      :      :
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
      :|||:      :      :      :      :      :      :      :      :
1565 ..ACAGTTGAAACCGAATTGATCTCTATCAATGGTAGCAATGATGAAGTG 1612
      :|||:      :      :      :      :      :      :      :      :
419 AspValAsnPhelValValGluGlnProSerGlySerThrIleAl 435
      |||||      :      :      :      :      :      :      :      :
1613 GATGTCGTATATAAGTCAAGAACGTACACGGGTAGTATCAACTTTGG 1662
      :|||:      :      :      :      :      :      :      :      :
435 alaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
      :|||:      :      :      :      :      :      :      :      :
1663 TATTGGTTACGGTCACAGAGTGGTATTAGTTATCAAGCAAGTGTCAAAC 1712
      :|||:      :      :      :      :      :      :      :      :
452 lAsnAsnPhelMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
      |||||      :      :      :      :      :      :      :      :
1713 AAGATAAATTTCTGGACACAGGGGGCGGCAAGTAAATAGTAGTGGTACGAAA 1762
      :|||:      :      :      :      :      :      :      :      :
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
      :|||:      :      :      :      :      :      :      :      :
1763 AATGATTATGGTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTAC 1812
      :|||:      :      :      :      :      :      :      :      :
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
      :|||:      :      :      :      :      :      :      :      :
1813 TAAAGATGGTGA.....AGCTTGGTGAAATGTTTCTTTGAAAACT 1856
      :|||:      :      :      :      :      :      :      :      :
502 yrvAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
      |||||      :      :      :      :      :      :      :      :
1857 ACGTAACCTCTAAAAGTGTATACATCTCTTAACATAAAGCGTACGACTTAT 1906
      :|||:      :      :      :      :      :      :      :      :
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
      |||||      :      :      :      :      :      :      :      :
1907 GGAAGTAATGTACTTTAGGTTTCCTCTGTAATGAAATAAATCACTCTATTA 1956
      :|||:      :      :      :      :      :      :      :      :
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
      :|||:      :      :      :      :      :      :      :      :
1957 TGTAGGATTAGCCCATACCTATATAATAAATAGTAACCTTTGCTCTAGAT 2006
      :|||:      :      :      :      :      :      :      :      :
544 lvclyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
      :|||:      :      :      :      :      :      :      :      :
2007 ATAACCGTAATTTATATATTCAATCAATGAAA..... 2038
      :|||:      :      :      :      :      :      :      :      :
561 GlyLysIleGlnValAspAsnGlnLysIleProAspPheLysHisAspTy 577
      :|||:      :      :      :      :      :      :      :      :
2039 .....TTTAAGGTAAATGGCATTAAACAAATGACTTTGATTT 2076
      :|||:      :      :      :      :      :      :      :      :
577 rThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspArgP 594
      :|||:      :      :      :      :      :      :      :      :
2077 TTCT.....TTTGGTTGGAACATAACACGCTTATAGAG 2111
      :|||:      :      :      :      :      :      :      :      :
594 rovalPheProThrGlnGlyMetSerHisSerValAsp.....Leuthr 608
      |||||      :      :      :      :      :      :      :      :
2112 GCTATTTCACCAACTAAGGGGTAAAGCAAGTCTTGGTGGCAGGAGTTACA 2161
      :|||:      :      :      :      :      :      :      :      :
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIl 625
      :|||:      :      :      :      :      :      :      :      :

```

```

2162 ATTCCAGGTTCTGTATAACAAATACTACAACTAAGTGCAGATGTACAGGG 2211
      :|||:      :      :      :      :      :      :      :      :
625 eTyrArgProPhe.....IleLysLysSerValL 635
      :|||:      :      :      :      :      :      :      :      :
2212 TTTCATACCATTAGACAGATACACCTCTGGGTTGTATCTGCAAAAGCAT 2261
      :|||:      :      :      :      :      :      :      :      :
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn....AsnLeuProPheTyr 650
      |||||      :      :      :      :      :      :      :      :
2262 CTGCAGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2308
      :|||:      :      :      :      :      :      :      :      :
651 GluAsnPhelTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      :|||:      :      :      :      :      :      :      :      :
2309 CAAACTATATACAGCGGTGTCATTTGATTTACCGGTTTTCCTTATGG 2358
      :|||:      :      :      :      :      :      :      :      :
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      :|||:      :      :      :      :      :      :      :      :
2359 TAGCAATTGGCCTAACGCA.....ATTATCAAGGTCAA 2393
      :|||:      :      :      :      :      :      :      :      :
684 lnrThrThrLeu.....GlyGluValValGlyGlyAsnAlaLeu 696
      :|||:      :      :      :      :      :      :      :      :
2394 ATAATAAATTTAATAAGATAAGTTCTGTATGTTGTTGTTATGCAATC 2443
      :|||:      :      :      :      :      :      :      :      :
697 AlaThrPheGlySerGlnLeuIleLeuProLeuProPheLysGlyAspTr 713
      |||||      :      :      :      :      :      :      :      :
2444 GCTACAGCTAGCGCAGAGTTAATTGCGCAACTCCATTTGTGAGTGATAA 2493
      :|||:      :      :      :      :      :      :      :      :
713 pIle...AspGlnValArgProValIlePheIleGluGlyGlyGlnValP 729
      :|||:      :      :      :      :      :      :      :      :
2494 GAGTCAAAATACAGTCCGAACCTCCCTATTGTTGATGCGGCAAGTGT 2543
      :|||:      :      :      :      :      :      :      :      :
729 heAspThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
      :|||:      :      :      :      :      :      :      :      :
2544 GGAATACTAATGGAATCAGATAAAATGGATTAGAGAGCAATGCTCTTG 2593
      :|||:      :      :      :      :      :      :      :      :
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
      :|||:      :      :      :      :      :      :      :      :
2594 AAAGAC.....TTACC 2604
      :|||:      :      :      :      :      :      :      :      :
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlat 778
      :|||:      :      :      :      :      :      :      :      :
2605 CGATTATGCAAAATCAAGCCGTACTCGCCCTCTACAGGTGTCGATTCC 2654
      :|||:      :      :      :      :      :      :      :      :
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
      |||||      :      :      :      :      :      :      :      :
2655 AATGGCAATCTCTCTAGTGACCAAGTGGTATTCTTCTTATGCTAAACCAAT 2704
      :|||:      :      :      :      :      :      :      :      :
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
      :|||:      :      :      :      :      :      :      :      :
2705 AAAAAATGAAAATGATGATGCGAACACAGTCCCAATTTAGTATTGGGGG 2754
      :|||:      :      :      :      :      :      :      :      :
811 rValPhe 813
      :|||:      :      :      :      :      :      :      :      :
2755 TTCTTTC 2761
      :|||:      :      :      :      :      :      :      :      :
seq_name: /SIBSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF60966
seq_documentation_block:
ID AAF60966 standard; DNA; 1830 BP.
XX
XX AAF60966;
XX
XX
XX 16-MAY-2001 (first entry)
XX
XX P. putida KT2440-associated DNA ORF00612.
XX
XX Transgenic plant; detection; probe; amplification; vaccine carrier;
XX microbial production strain; biological remediation; ds.
XX
XX Pseudomonas putida.
XX
XX DE19935088-A1.
XX
XX 01-FEB-2001.
XX
XX PD

```

XX 27-JUL-1999; 99DE-1035088.
 XX
 XX
 PR 27-JUL-1999; 99DE-1035088.
 XX
 PA (TIGR-) TIGR INST GENOMIC RES.
 PA (QUIA-) QUIAGEN GMBH.
 PA (GBFB) GES BIOLOGISCHE FORSCHUNG MBH.
 PA (DKFZ) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
 PA (MED1-) MEDIZINISCHE HOCHSCHULE HANNOVER.
 XX
 DR WPI; 2001-192469/20.
 XX
 PT New DNA sequences specific for pseudomonas putida KT2440, useful as
 PT safe genetic engineering host, allow detection in presence of other
 PT related bacteria
 XX
 PS Claim la; Page 13-14; 158pp; German.
 XX
 CC This invention describes novel DNA sequences (I) for specific detection
 CC of Pseudomonas putida KT2440. The invention also describes (1)
 CC recombinant expression vector containing (1); (2) prokaryotic or
 CC eukaryotic cells transformed or transfected with (1) or the vector of
 CC (1); (3) production of expression products by culturing cells of (2);
 CC (4) expression products, or their fragments, of (1) and synthetic
 CC proteins or peptides with the same sequences (A); (5) poly- or
 CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
 CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
 CC plants that contain transformed or transfected cells of (2); (8)
 CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
 CC carrying one or more (1). (1), and their fragments, are used as probes
 CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
 CC polymerase chain reaction, and for production of transgenic plants, (1),
 CC or antibodies that recognize their expression products, are used for
 CC detecting the presence of KT2440, particularly in presence of other,
 CC even closely related, bacteria. KT2440 is one of the bacteria classified
 CC as safe, by the National Institutes of Health, for genetic engineering
 CC work, e.g. as microbial production strains, for biological remediation
 CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
 CC homology with sequences in other bacteria (specifically the closely
 CC related pathogen P. aeruginosa). Compared with other 'safe' bacteria, it
 CC has greater catabolic activity and better survival in, and adaptation to,
 CC the rhizosphere and soil.
 XX
 SQ Sequence 1830 BP; 401 A; 555 C; 529 G; 344 T; 1 other;

alignment_scores:
 Quality: 927.00 Length: 502
 Ratio: 2.459 Gaps: 3
 Percent Similarity: 75.100 Percent Identity: 37.251

alignment_block:

US-09-701-711-2 x AAF60966 ..

Align seg 1/1 to: AAF60966 from: 1 to: 1830

16 MetaAlaValMetValMetSerThrHisAlaGlnAlaAlaAspPheMe 32
 |||
 28 ATGTCGGCAGCATGATCGCTGAAGTTCACGCCGAGTCC.....TTCAC 71
 32 talaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleGluSerL 49
 |||
 72 CATCTCCGATATTCGCTCAACGGGCTGCAGCGGTTTCGGTGGCAGTG 121
 49 euGlnSerValLeuProPheArgLeuGlyGlnValValSerGluAsnGln 65
 |||
 122 TGTTCGGTGCCTTCGCTGAACGTCGCGAGCCAGGCTGACGACCCGCG 171
 66 LeuAlaAspGlyValLysAlaLeuThrAlaThrGlyAsnPheSerAspVa 82
 |||
 172 CTGGTGGACTCGACTCGTTCCTCTCAAGACCGGGTCTTCCAGGACAT 221

82 lGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValThrGluArgp 99
 |||
 222 CCAGTTGACCGCGGATGGCAATGTGCTGATCATCAACGTTGGTGCAGCGCC 271
 99 roLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGlu 115
 |||
 272 CGTCGGTGTGAGCATCGAGATTGAAGGCAACAAGGCGATCAGCACCGAA 321
 116 GlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLe 132
 |||
 322 GACCTGATGAAGGCGCTAAGCAATCGGCGCTGCCGAAGTGGTGGCAG 371
 132 uLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrI 149
 |||
 372 CCAGCGTGCACCTCGAAGGTGTGCTAAGCACTGCAACGCCAGTACG 421
 149 leSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnThrMetLeu 165
 |||
 422 TGGCCCAAGGCGCTACTCGGCGAGTGCATGCCGAAGTGGTGGCAG 471
 166 AspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyLysProAl 182
 |||
 472 CCGCGCAACCGTGTGGCCCTGAAGATCAAGATCAACGAAGGCCACCGTCG 521
 182 aArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspAlaA 199
 |||
 522 CCGCATCCAGCACATCAACATCGTTGGCAACAACGTTATTCGATGATGAGA 571
 199 sPheLeuAspValLeuAlaIleLys...AspAsnLysIleAsnProLeu 214
 |||
 572 CCGTCGGGCGAGTGTTCGAGCTGAGACACCAACCACTGGCTGCTGCTTC 621
 215 SerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAs 231
 |||
 622 AGAAGCAGCAGCAAGTACCCCGTGAAACACCTCCGCTGACCTGGAGCG 671
 231 nLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysA 248
 |||
 672 CCGTGTCTCTACTCTGAGCGCGCTACATCAACATGGACATCGCCT 721
 248 sPAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValGluIle 264
 |||
 722 CCACCCAGGTGTCCATCACCGCGCAAGACGACGCTACATCACCCGCT 771
 265 SerLeuHisGluGlyGlnTyrArgPheGlyGlnThrGlnPheLeuG 281
 |||
 772 AACATCAACGAAGCGGAGAAATACACCGTTCGCGACGTGAAGCTGTCGG 821
 281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLysPheLysA 298
 |||
 822 TGACCTCAAGTGCAGGAGACGAGTCAAGTCAAGTCTGCTGCTGGTGCAGC 871
 298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314
 |||
 872 CCGGGCAGGTATTCGCGCAAGGTGATGACCCACGCTCCGAGGTGATC 921
 315 SerThrLysPheGlyAspAspGlyTyrTyrAlaGlnIleArgProVa 331
 |||
 922 ACCCGCGCGCTGGTAACGAGGTACACCTTCGCTACGCTCAACGCGCT 971
 331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348
 |||
 972 GCCGCAACCAACGACGAGGACACGAGTCAAGTCAAGTCTGCTGCTGGTGC 1021
 348 sProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364
 |||
 1022 ACCCGGGCAAGCGCTGCTACGTCACCGCATCACTACCGCGGCAACACC 1071
 365 LysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381
 |||
 1072 AGACCGAAGACGAGTGTGCTGCGGAATGCGGCGACATGGAAGCGCG 1121
 381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398

; |||||::: |||::: |||::: |||||::: |||
 1122 CTGGCGGTGACCTACTGTATGCACGATCCAGACCCGTTGAGCGCC 1171
 398 hrGlyPhePheLysHisValThrValAspThrArgProValProAsnSer 414
 |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
 1172 TGGGCTTCTTCAAGGAAGTCAAGTCGAGACCCCGAGGTGCCTGGCACT 1221
 415 ProAspGlnValAspValAsnPheValValGluGlnProSerGlyse 431
 |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
 1222 GACGACCAAGTCGAGCTCAACTACAGGCTCGAAGACGAGCGCTCGGCCT 1271
 431 rSerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnP 448
 |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
 1272 GATCACCGCCAGCGTGGGTTCGCCAGAGCGCGGCTGATCCTGGGCT 1321
 448 heAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAla 464
 |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
 1322 GTTCGATCAGCCAGACCAACTCTCTCGGTACCGGTACACAGGTATCCATC 1371
 465 SerPheSerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAs 481
 |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
 1372 GGCTGTGACCCGTTCGGATACACACCGGTACAACTTCGGCTTCGTTGA 1421
 481 pProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrA 498
 |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
 1422 TCCCTACTTCACGGCGATGGCGTCTCCACAGATNGTT..... 1458
 498 rgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSer 514
 |||||::: |||||::: |||||::: |||||::: |||||::: |||||:::
 1459GACAGCCCACTCTTGAAAAACACAGGATTCGCACAAT 1494
 515 TyrGly 516 .
 ||||
 1495 CTTGGA 1500

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:AAS81841

seq_documentation_block:

seq_documentation_block.
ID AAS81841 standard; CDNA; 2057 BP.

XX
AC

DE DNA encoding novel human diagnostic protein #17645,

XX

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AA594564 represent novel human diagnostic coding sequences of the invention.

Sequence 2057 BP; 505 A; 495 C; 550 G; 507 T; 0 other;

alignment_scores:

Quality:	901.00	Length:	648
----------	--------	---------	-----

Ratio:	2.110	Gaps:	9
--------	-------	-------	---

Similarity: 65.895 Percent Identity: 33.488

alignment_block:

US-09-701-711-2 x AAS81841

Align seq 1/1 to: AAS81841 from: 1 to: 2057

14 MetThrMetAlaValMetMetValMetSerThrHisAlaGlnAlaAlaAs 30

```

::|||::: ::: ||| ::::: |||:::
::|||::: ::: ||| ::::: |||:::
::|||::: ::: ||| ::::: |||:::
::|||::: ::: ||| ::::: |||:::

```

157 TTGCTCATAGCGTCGCTGCTGCTTAGCAGCGCCACCGTATACGGTGCTGA 206

30 p...PheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrI 46

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

207 AGGTTCGTAGTGAAGATATTCATTTTCAAGGCCCTTCAGCGTGTCGCCG 256

16 [0C]UGRTauGIngoVzJTanDnrbtbnrtcuClwVzJban

46 IeGLuSerLeuGlnSerValLeuPropheArgLeuGlyClnValValSer 62

257 TTGGTGGGGCCCTCCTCAGTATGCCCGCCACAGCCGACACCGCTTAT 306

ZZJ/ 1TGGTGGGCCCCICAGTAHGCCGGTGGCACAGGCACACGGTFAAT 306

63 GluAsnGlnLeuAlaAspGlvvalIysAlaLeuTyrAlaThrGlyAsnph 79

00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99

307 GATGAAGATATCAGTAATACCATTCGCGCTCTGTTTGCTACCGGCAACTT 356

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

79 eSerAspValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValT 96

[illegible]

357 TGAGGATGTTCCGGTCCCTTCGTGATGGTGATACCCCTTCTGTGTCCTCGTGC 406


```

1171 GGGCCCTATTCCACGCTGGCAGGCTTACTATTCCGACGCGAC..... 1215
683 nGlnThrThrLeuGlyGlnValValGlyGlyAsnAlaLeuAlaThrPheG 700
1216 .....ACCGATGTGCTGGGGCAATATCTCGTGAAGGACAG...AGCCA 1299
700 lySerGluLeuLeuLeuProLeuProPheLysGlyAspTrpIleAspGln 716
1253 GTCCGAGTACCTGTTCGCCCTCCCTCTGTGAAGGACAG...AGCCA 1299
717 valArgProvalIlePheLeuGlyGlyGlnValPhe...AspThr.. 731
1300 CTGCGCAGCTCGGTGTCTGCGATGCGGCACGACGTATGCGCAGCCTG 1349
732 .....ThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheL 745
1350 CTACCTGTCCACCCAGCGGTCGCGCAGCGTCGACCTGGCG..... 1392
745 ysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPro 761
1392 ..... 1392
762 LeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThr 778
1393 .....CAGATGGCTGTGCTGCTGGGGTTCGCGTGAC 1424
778 rTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
1425 CTGGTACAGCCCGGCGGCTGAGTTTCAGCCCTGGCGGCCCATTA 1473

```

seq_name: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:AAI97964

seq_documentation_block:

ID AAI97964 standard; DNA; 6617 BP.

AC AAI97964;

DT 20-NOV-2001 (first entry)

DE Lawsonia intracellularis coding sequence SEQ ID NO: 1.

KW HtrA; PonA; HypC; YefW; ABCI; Omp100; Lawsonia intracellularis infection; vaccine; ds.

OS Lawsonia intracellularis.

PN JP20001169787-A.

PD 26-JUN-2001.

PF 20-OCT-2000; 2000JP-0320736.

PR 22-OCT-1999; 99US-0160922.

PA (PFIZ) PFIZER PROD INC.

DR WPI; 2001-592540/67.

XX Lawsonia intracellularis polynucleotide and encoded protein, used to prevent Lawsonia intracellularis infection -

PS Claim 3; Page 37-39; 67pp; Japanese.

CC The present invention provides isolated polynucleotides encoding HtrA, PonA, HypC, lysS, YefW, ABCI or Omp100 protein of Lawsonia intracellularis. The sequences can be used in vaccines for the prevention of Lawsonia intracellularis infection. The present sequence is a coding sequence of the invention.

SQ Sequence 6617 BP; 2216 A; 1153 C; 1142 G; 2106 T; 0 other;

```

alignment_scores:
  Quality: 569.00      Length: 824
  Ratio: 1.284        Gaps: 27
  Percent Similarity: 53.762      Percent Identity: 23.908

alignment_block:
US-09-701-711-2 x AAI97964 ..
Align seg 1/1 to: AAI97964 from: 1 to: 6617

35 AspileThrIleThrGlyLeuGlnArgValThrIleGluSerLeuGlnSe 51
   ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4190 GATGTAGTATTTCATGGCTTAAAGTTCTTGATCTGATGTAATCTTAC 4239

51 rValLeuProPheArgLeuGlyGlnValSerGluAsnGlnLeuAla 68
   : ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4240 ACCACCTACTATTAAAGGAGATCATCATGATGCCCAAAATTAATG 4289

68 spGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnVal 84
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4290 CAGAAATCAAAAAAATATGGAATTAGGATATTTAGTGATGCTCTGCA 4339

85 TyrHisGln.....GluGlyArg...IleIleTyrGlnValThrG 97
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4340 AGTATTGAAGAAACGGGGAAGGACGATTACTTGTATTACTGTACAAGA 4389

97 uArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProL 114
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4390 AAAGCCTAAAAATTACAGATGTTGTTCAAGGCTCAAAAGCTGTAAAGTA 4439

114 ysGluGlyLeuGlnGluGlyLeuLysAsn...AlaGlyLeuAlaValGly 129
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4440 TCATAACATCTTGCCTGCAATGAGTTCTAAAAAAGGATCAGTTATTAGT 4489

130 GlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAs 146
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4490 GATGACTATTGTCCCAAGATATTCAA.....AAATATCCGA 4527

146 nGlnTyrIleSerGlnGlyTyrTyr.....AsnThrGluIleT 159
   : ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4528 CCTCTATAGAAAGAGGCTACTATCTCGTGAAGTTAATATGAATAAA 4577

159 hrValLysGlnThrMetLeuAspLysAsnArgValLysLeuAspMetThr 175
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4578 AACAGAAAGAA.....AATACTTCTTCGCAACACTATTGTTAACA 4618

176 PheAlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAs 192
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
4619 GTAAATGAGGGAAAAAATTTATATTAAAGATGTCGGAATTGAAGGACT 4668

192 nGlnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAsp. 208
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4669 TGAACAATAAAGCTTAAACTTTTAAAAAAGAGTTAGCATTAACAGAAC 4718

209 ..AsnLysIleAsnProLeuSerLysAlaAspArgTyrThrGlnGluLys 224
   ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4719 GTAATTTTTCATCGTTTACTTGGACACAGGTGTATTACGTGAAGATAT 4768

225 LeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPhe 241
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4769 CTTGAACGTGACTCTCTAGCAATCTCTGCCTATGCCATGAATCATGGCTA 4818

241 eValArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysA 258
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4819 TGTAGATATTCAAGTTGCTTCACTGAGTAAACATTCATCAATGAA.....A 4862

258 snArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPhe 274
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4863 AAGGAATGTTATTACATTTAGAGTAAAGAGTAAGCGCTATAAATA 4912

275 GlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuG 291
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
4913 GGAATAATAGACTTTAAAGGAGATCTTATTGAGACAATGAACAACCTCT 4962

```

291 uAlaLeuLeuLysPhe.....LysAlaGluGluGlyPheSerG 304
::: |||:::
4963 TAAAGTAAACAAATTTGATGATCATAAAACTATGAGCAGATTTTCTC 5012
304 InAlaMeLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
:::||||:::
5013 TTTCTGTTATGCAAGATGATGTAAGAGCATTAACAGATTTTATTTCAGAT 5062
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG1 337
|||||:::||||:::
5063 TATGTTATGATTTGCTGAGAGTAGATCTTGAACAACCAACCAAAATGAAGA 5112
337 userArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
|||:::||||:::
5113 AGATCAACAATTTGATGTTACTTTCTTCTTATTGATAAAAAACAAGTCT 5162
354 yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
:::|||||:::
5163 TTCTTCGTAGAATAATTTGTTGAAGGAATACTCGTACTAGAGATAATGTT 5212
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
:::|||||:::
5213 ATCTCCGTGAATTCAGCTTCGTGATGAGATCTTTTAAATGGTCAACA 5262
387 sileGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
:::|||||:::
5263 TCTCCGACGCTCTAATGAATGCTTAAACGCTTGGCTTATTTAAACCAAG 5312
404 alThrValAspThrArgProValProAsnSerProAspGlnValAspVal 420
:::|||||:::
5313 TAGATACAGATACACCTGCTACA...GGGAAGATGATCAAGTTGATCTA 5359
421 AsnPheValValGluGluGlnProSerGlySerSerThrIleAlaAlaG1 437
|||||:::||||:::
5360 CTGTGTAAGTTCAAGAGCTCGACAGGTGCAATCAGGTGGTGGTGG 5409
437 yTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsnA 454
|||||:::
5410 TTACTCAACACATTTCTAAATTTGGTGTTCAGGAAGTATCTCAGAAAGAA 5459
454 snPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArgSerGlu 470
|||:::||||:::
5460 ACTTATGGGAAAGGTATATTTTAAGTATGAAGTTTATTTCTAGT 5509
471 ThrArgGluValTyrSerLeuGlyMetThrAsnPro.....TyrPheTh 485
:::|||||:::
5510 AAGTCATCTCTCTTGTGATCTTTCTTTTACCAATCCTCGTGTATTATGATAC 5559
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
| ||| |||:::||||:::
5560 AGACTTTGGCTTTAGTATAACATT.....TATACGCTACGAGATGAAT 5603
502 yrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySer 518
:::|||||:::
5604 GGGAT.....GACITTCGGTAAAAAACTTATGGAGATACC 5638
519 LeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLe 535
:::|||||:::
5639 ATACGCTATTTCACCCCTTAGGAGAATATTCATCTATCTTTGTTGGCTA 5688
535 uAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMetGlyIleSera 552
:::|||||:::
5689 TCGAATTGATCAATATCGTCTATAT..... 5713
552 snValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnAsn 568
:::|||||:::
5714GATATTCATCTACAGCACCGCTCTTATCTTACACTATCA 5755
569 GlyIleProAspPheLysHisAspTyrThrTyrAsnAlaIleLeuG1 585
|||:::||||:::
5756 GGG.....AAAAATTTCTAGTGTAGTAAGTGGT.....GG 5787

585 yTrrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMets 602
:::|||||:::
5788 TTTTACTTTTGTATTCTACACAGCTCGTGAGAGACCATCTAAAGG..... 5833
602 erHisSerValAspLeuThrValGlyPheGlyAspLysThrHisGlnLys 618
|||:::||||:::
5834 ..CATATTGCCAAACTAATTTGTTGAATATGGAGGT.....GGTGGT 5872
619 ValValTyrGlnGlyAsnIleTyrArgProPheIleLysLysSerValLe 635
:::|||||:::
5873 CTGTGGTGAATGATAACTTCTTCAAGCCCAATTTGCTGAA.....CT 5913
635 uArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsn..... 646
||:||||:::
5914 ACAAGGATTTTACTCAATTTCAAGAAGTAAACCAACCATATAATACATTGCG 5963
647LeuProPhe 649
5964 GTACACGTGCGAGTGCAGCTTATAACAATAGTAAAAAACCTGTGCCAGTA 6013
650 TyrGluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspG1 666
:::|||||:::
6014 TTTGACCGATTTTATTTGGTGTATAGATAGTATTAGAGGATATGATAC 6063
666 nSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyG 683
:::|||||:::
6064 AGAAGATCTTGCAACAAAGATCTCTCGCTT..... 6094
683 InGlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAla..... 697
|||:::||||:::
6095GGAGATGAATTTGGTGGTAGTAGATGGCTTTCTTCTT 6130
698ThrPheGlySerGluLeuIleLeuProLeuPr 708
|||||:::
6131 RACCTAGAGTATATTGGACATTCCAGCCAGAGTAGGTCTTGCAATTA.. 6178
708 oPheLysGlyAspTrrPileAspGlnValArgProValIlePheIleGluG 725
||| |||:::
6179GTTCCATTCTATGACA 6194
725 LyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeu 741
||| |||:::||||:::
6195 TAGGA.....TTCCAACA.....GATTCT 6214
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
:::|||||:::
6215 GTACAAACTTCTTAACCCATCTCT..... 6238
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
:::|||||:::
6239AACTCAACCAATCATATGGCC 6260
775 aGlyAlaThrTrrTyrThrProIleGlyProLeuSerIleSerTyrAla 791
:::|||||:::
6261 TTGAACCTTCGCTGGCTTCCACCAATGGGAGATTTCGATTTGCCTATGGT 6310
792 LysProLeuAsnLysLysGlnAsnAspGln.....ThrAspThrValG1 806
||| ||| |||:::
6311 ATACCACCTCAATAAAAAATGTTAGTGGCAAAAAAACTCGTGGTAGATTGA 6360
806 nPheGlnIleGlySerValPhe 813
||:||||:::
6361 ATTTTCAATGGGCAATCTTTC 6382

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAF61056

seq_documentation_block:

ID AAF61056 standard; DNA; 1602 BP.

XX

AAF61056;

XX

16-MAY-2001 (first entry)

XX

[illegible]

1938 ACCAGAAATAGAGGTATTTGGGATAGGGATTAC...CATACGCCCTATCA 1984
579 hrTyraAsnAlaIleLeuGlyTrpAsnTrpSerSerLeuAspArgProVal 595
1985 CCAGCTCTTTCACCTTGATGTGAGTATGACACACCGATGATTATTAC 2034
596 PheProThrGlnGlyMetSerHisSerValAspLeuThrVal..... 609
2035 TTCCCTAGAAATGGGGTTATCTTTAGTTCCCTATGCGAGCATGCTGGCTT 2084
610GlyPheGlyAspLysThrH 616
2085 GCCAAGCTCTGGCAGCTCAATCTTGAACGGTTAGCGGGAATGTCC 2134
616 isGlnLysValValTyr.....GlnGlyAsnIleTyrArg 627
2135 GTAACACCAAGTTTATGGTAAATTCGCCGCTTACCACCATTTGCCAAAA 2184
628 ProPheIleLysLysSerValLeuArgGlyTyrAlaLysLeuGlyTyr.. 643
2185 TATTATTATGATAGATTGATCGCTCGCTTTAAACCGAAGGAGTTATAT 2234
644GlyAsnAsnLeuProPheTyrGluAsnPheTyrA 655
2235 CTTTATAGTATAACACCGATGATTACTTCGCCCTTAAACTCCACCTTCTACA 2284
655 laGlyIleTyrGlySerValArgGlyTyrAspGlnSerSerLeuGlyPro 671
2285 TGGGGGCGTAACCCGGTAGAGGCTTTAGAACGGATCGGTACTCCCT 2334
672 ArgSerGlnAlaTyrLeuThrAlaArgGlyGlnGlnThrThrLeuGln 688
2335 AAAGATGAG.....TTTGG 2348
688 yGluValValIcGlyAsnAlaLeuAlaThrPheGlySerGluLeu.... 703
2349 CTTTGTGGCTTGGAGCGCATGGGATTTTACCCTCTTCTACTGAATTGAGCT 2398
704IleLeuProLeuProPheLysGlyAspTrpIleAspGlnValArg 718
2399 ATGGGTGCTAAAGCGCGCT.....AAATGCGC 2427
719 ProValIlePheIleGluGlyGlycIcIcValPheAspThrThrGlyMetAs 735
2428 TTAGCGGTGTTTTTTTGACTTTGGT..... 2451
735 pLysGlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaG 752
2452TTCTTAACTTTAAACCCCA.....A 2473
752 luGlnAsnAlaLysAlaAlaArgProLeuLeuThr..... 764
2474 CTAGAGGGAGTTTTTCTATTAACGCTCTGTGTACACACCGAATTTTAA 2523
765GlnAspLysGlnLeuArgTyrSe 772
2524 GATTATGGCGTTATAGGGCTGGGTTTGAAGAGCGACTTGGAGGCGCTC 2573
772 rAlaGlyValIcIcAlaThrTrpTyrThrProIleGlyProLeuSerIles 789
2574 CACAGGCTTCAGATTGAATGGATTTCGCCATGGGCGCTTTGGTGTGA 2623
789 ertyr..... 790
2624 TTTTCCCTATAGCGTTTTTCAACCAATGGGGGATGGCAATGCCAAGAA 2673
791 AlaLysProLeu.....AsnLysLysGlnAsnAspGlnThrAspThrVa 805
2674 TGTAAAGGGCTATGCTTCAACCCCTAACATGGACGATTACACGCAACACT 2723
805 lGlnPheGlnIleGlySerValPhe 813

230 uAsnLeuArgAlaLysTyrZeuAsnaLa.....GlyPheValA 243
 ::::| | | | | | | | | | | | | | | | | | | | | |
 852 GTATATGCCTAGGGGTACTTAGACGCTCATATTTCTTCGCCCTTTTTTGA 901
 ::::| | | | | | | | | | | | | | | | | | | | | |
 243 iq.....PheGluIleLysAspAlaLysLeuAsnIleAsncluAsPlys 257
 ::::| | | | | | | | | | | | | | | | | | | | | |
 902 AAACGGATTTCACCACCCATGACGCTAACGCTCCT..... 936
 ::::| | | | | | | | | | | | | | | | | | | | | |
 258 AsnArgIlePheValcluflesSerLeuHisGluGlyGluGlnTyrArgPh 274
 ::::| | | | | | | | | | | | | | | | | | | | | |
 937TATAAGGTCAAAGAGGGGATCCAATACAGGAT 968
 ::::| | | | | | | | | | | | | | | | | | | | | |
 274 eGlyGlnThrGlnPhe...LeuGlyAsnLeuThrTyrThrGlnAlaGluL 290
 ::::| | | | | | | | | | | | | | | | | | | | | |
 969 TTCAGATATTTTAATAGAGATTGACAACCGGTAGTCCCTTAAAAACCT 1018
 ::::| | | | | | | | | | | | | | | | | | | | | |
 290 euGluAlaLeuLeuLysPheLysAlaGluGlyPheSerGlnAlaMet 306
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1019 TAGAAAACCGCTTAAGTTAAAGGAAGAAGATGCTTTAATATTGAGCAT 1068
 ::::| | | | | | | | | | | | | | | | | | | | | |
 307 LeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTy 323
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1069 TTAAGAGCGGATGCCAAATTTAAAAACCGAAATCGCGATAGGGCTA 1118
 ::::| | | | | | | | | | | | | | | | | | | | | |
 323 rTyrTrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgT 340
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1119 TCGGTTTGCGGTGGTCAAGCCAGACTTGGATAAAGACGAAAAAACGGCG 1168
 ::::| | | | | | | | | | | | | | | | | | | | | |
 340 hrValAspValGluTyrThrIleAspProValHisProValTyrValArg 356
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1169 TTGTGAAGTCATTTATCGATTGAAGTGGCGATPATGGTCATATCAAT 1218
 ::::| | | | | | | | | | | | | | | | | | | | | |
 357 ArgIleAsnPheThrGlyAsnPhelysthrGlnAspGluValLeuArgAr 373
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1219 GATGTCATATTTCAAGGNACCGAGCGCATAGGATAGGATAGGAG 1268
 ::::| | | | | | | | | | | | | | | | | | | | | |
 373 gLUmetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnL 390
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1269 GGAATTG.....TTACTAGGGCCTAAAGATAAATACAAC 1303
 ::::| | | | | | | | | | | | | | | | | | | | | |
 390 euSerArgAlaArg.....LeuMetArgThrGlyPhePhe 401
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1304 TGACCAAATCGAATAATCCGAAAAATCTTTGAGCGCTTAGGGTTTTTC 1353
 ::::| | | | | | | | | | | | | | | | | | | | | |
 402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1354 TCATAAGTCACAGATTGAGAAAAAAGGTCAATAGCTCA.....TTGAT 1397
 ::::| | | | | | | | | | | | | | | | | | | | | |
 418 lAspValAsnPheValValGluGlnProSerGlySerSerThrIleA 435
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1398 GGATTTGTTAGTGAGCGTAGAGAGGGCGCACCGGGCAGTTGCCAATTCG 1447
 ::::| | | | | | | | | | | | | | | | | | | | | |
 435 laalagLyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSer 451
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1448 GGTTCGGCTATGGCTCTTAGGAGGGCTCATGCTTAATGGGAGCGTGAGC 1497
 ::::| | | | | | | | | | | | | | | | | | | | | |
 452 GlnAsnAsnPheMetGlyThrGlyLysHisValasn.....AlaSerPh 466
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1498 GAAGAGATCTTTTGGCACAGGGCAAGCATGAGCTTGATGTATCAAT 1547
 ::::| | | | | | | | | | | | | | | | | | | | | |
 466 eSerArgSerGluThrArgGlu..... 473
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1548 TGCCACAGGGGGGTAGATCTTATCCGGGCATGCCAAAGGGCGGGC 1597
 ::::| | | | | | | | | | | | | | | | | | | | | |
 474ValTyrSerLeuGlyMetThrAsnProTyrPheThrValasn 487
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1598 GTATGTTTCGGGAATTTGAGCTTGACTAATCCAAGGATTTTGCACAGC 1647
 ::::| | | | | | | | | | | | | | | | | | | | | |
 488 GlyValSerGlnSerLeuSerGlyTyrThrArgLysThrLysTyrAspAs 504
 ::::| | | | | | | | | | | | | | | | | | | | | |
 1648 TGGTATAGCTCTACGATCAATCTTTAT.....GCGGATTA 1682
 ::::| | | | | | | | | | | | | | | | | | | | | |

```

2479 .....TTCTTAACCTTTAAACCCCA.....ACTAGGGGA 2509
      :::: ||||| ||| :::::
755 laLysAlaAlaAsnArgProLeuLeuThr..... 764
      :: ||| ||| |||
2510 GTTCTCTCTATACGCTCCACACGCGCGAATTTAAAGATTATGGC 2559
765 .....GlnAspLysGlnLeuArgTyrSerAlaGlyVa 775
      :::: ||| ||| ||| |||
2560 GTTGTAGGGCTGGTGTGAAGGGCGACTTGGAGGCTCTACAGGCTT 2609
775 lGlyAlaThrTyrThrProLleGlyProLeuSerIleSerTyr.... 790
      ||| ::::: ||||| ||| ::::
2610 ACAGATTGAATGATTCGCCCATGGGCTTGTGTTGATTTCCTCA 2659
791 .....AlaLysPro 793
2660 TACCGTTTTCACCAATGGGCGATGGCAATGGCAAAATGTAAGG 2709
794 Leu.....AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG 808
      ||| ||| ::::: ||||| ||| ::::
2710 CTGTGCTTTAACCTAACATGACGATTACACGCAACATTTTGAATTTC 2759
808 nileGlySerValPhe 813
      ::::: ||| |||
2760 TATGGGAACAAGTTT 2775

```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-embl/NA1999.DAT:AAx75802

seq_documentation_block:
ID AAx75802 standard; DNA; 2778 BP.

XX AC AAx75802;

XX DT 03-AUG-1999 (first entry)

XX DE H. pylori outer membrane polypeptide encoding DNA.

XX KW Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
XX KW cellular immune response; ds.

XX OS Helicobacter pylori.

XX PN W09921959-A2.

XX PD 06-MAY-1999.

XX PF 28-OCT-1998; 98WO-US22883.

XX PR 17-DEC-1997; 97US-0993001.

XX PR 28-OCT-1997; 97US-0999131.

XX PA (GENO-) GENOME THERAPEUTICS CORP.

XX PI Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;

XX DR WPI; 1999-326698/27.

XX DR P-PSDB; AAY17183.

XX PT Cellular vaccine against Helicobacter pylori

XX PS Claim 8; Page 147-149; 352pp; English.

XX CC The invention relates to a vaccine for preventing or treating infections
XX CC by Helicobacter pylori. The vaccine contains at least one isolated
XX CC H. pylori polypeptide, or its fragments, in a carrier, where the
XX CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a
XX CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce
XX CC humoral and cellular immune responses. The vaccines are used to treat or
XX CC prevent infections by H. pylori. Sequences AAx75779 to AAx75837 represent
XX CC nucleic acid sequences encoding H. pylori outer membrane polypeptides
XX CC (OMPs) AAY17160 to AAY17218.

SQ Sequence 2778 BP; 813 A; 514 C; 681 G; 770 T; 0 other;

alignment_scores:
Quality: 429.00 Length: 922
Ratio: 0.943 Gaps: 32
Percent Similarity: 49.349 Percent Identity: 21.475

alignment_block:
US-09-701-711-2 x AAx75802 ..

Align seg 1/1 to: AAx75802 from: 1 to: 2778

```

30 AspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIl 46
      ::::: ||||| ||| :::::
232 GAATGAAGTCAAGTCCATTTCTTATGCGGGCTTCTTACATGCTGA 281
46 eGluSerLeuGlnSerValLeuProPheArgLeuGlyGlnValValSerG 63
      ::::: ||||| ||| :::::
282 CATGCTCGCTAATGAAATTCGCGTGGCGATATGCTGGATT 331
63 luAsnGlnLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPhe 79
      ::::: ||||| ||| :::::
332 CTAATAAATAGACCGCTGTTTACCTTTGTCAACCAAGGGTATTT 381
80 SerAspValGlnValTyrHisGlnGlyArgIleIleTyrGlnValTh 96
      ::::: ||||| ||| :::::
382 AAAGAGCTTTATGCCACTTTTGAACACGCGATTTTAGAGTTTCATTTGA 431
96 rGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleP 113
      ::::: ||||| ||| :::::
432 TGAATAAGCCAGGATTCGCGGGGTAGAAATCAAGGT.....TATG 472
113 roLysGluGlyLeuGlnGlyLeuLysAsn....AlaGlyLeuAlaVal 128
      ::::: ||||| ||| :::::
473 GGACTGAAAGGAAAGACGCTTAAATCCCAATGGGATCAAAAG 522
129 GlyGlnProLysGlnAlaThrValGlnMetIleGluThrGluLeuTh 145
      ::::: ||||| ||| :::::
523 GCGACACCTTTTATGAGCAAAATTTAGAGCATGCTAAAACGCTTTAAA 572
145 rAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysG 162
      ::::: ||||| ||| :::::
573 AACGGCTTTAGAGGGCGAGGCTATTATGGGCGGTGGAGGTCGCCA 622
162 lNThrMetLeuAsp..... 166
      ::::: ||||| ||| :::::
623 CAGAAAGTCACTGAGGAGCGCTTATTGATCGTTTGTGTAATAGG 672
167 GlyAsnArgValLysLeuAspMetThrPheAlaGluGly.....LysPr 181
      ::::: ||||| ||| :::::
673 GGGACAGTATTATATCAACAATCCATTATGAGGAAGCGATAAAT 722
181 oAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspA 198
      ::::: ||||| ||| :::::
723 AAAACGCGCTGTGATGTAATCTTTGAGCGCGAACAAG.....CAGC 763
198 laAspLeuIleAsp...ValLeuAlaIleLysAspAsnLysIleAsnPro 213
      ::::: ||||| ||| :::::
764 GCGATTTTCATGGCTGGATGGGCTTGAATGACGGGAATTG..... 807
214 LeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGl 230
      ::::: ||||| ||| :::::
808 .....CGCTTAGATCAATTAGAATACGATCTTTTTCGATCAAGATGT 851
230 uAsnLeuArgAlaLysTyrLeuAsnAla.....GlyPheValA 243
      ::::: ||||| ||| :::::
852 GTATATGCGTGGGCTTACTTAGACGCTCATATTTCTGCGCTTTTGA 901
243 tq.....PheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLys 257
      ::::: ||||| ||| :::::
902 AAACGGATTTTCCACCATGACGCTAAGCTCCAT..... 936

```


2560 GTGTAGGGCTGGGTTTGAAGGGCGGACTTGGAGGGCTTCTACAGGCTT 2609
775 lglYAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyr.... 790
2610 ACAGATTGAATGATTCGCCCATCGGGCTTGGTGTGATTTCCTTA 2659
791AlaLysPro 793
2660 TAGCGTTTTCACCAATGGGCGATGGCAATGGCAAAAATGTAAGGG 2709
794 leu.....AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGln 808
2710 CNGTCTTTAACCCCTAACATGACGATACACGCAACATTTGAAATTTTC 2759
808 nleGlySerValPhe 813
2760 TATGGGACACAGGTTT 2775

seq_name: /SIDS1/gcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAF91443

seq_documentation_block:

ID AAF91443 standard; DNA; 1001 BP.

XX AC AAF91443;
XX
XX 04-MAY-2001 (first entry)
XX
XX Haemophilus influenzae (H1rD) Omp26 gene upstream sequence, SEQ ID:69.
XX
XX Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
XX genetically modified; protective antigen expression; LPS detoxification;
XX LPS; lipid A; homologous recombination vector; immunisation;
XX immunoprotective; non-toxic; paediatric; ds.
XX
XX Haemophilus influenzae.
XX
XX W0200109350-A2.
XX
XX 08-FEB-2001.
XX
XX 31-JUL-2000; 2000WO-EP07424.
XX
XX 03-AUG-1999; 99GB-0018319.
XX
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Berthet FJ, Dalemans WLJ, Denoel P, Dequesne G, Feron C, Lobet Y;
XX Poolman J, Thiry G, Thonnard J, Voet P;
XX WPI; 2001-138654/14.
XX
XX New isolated polynucleotide useful for outer membrane vesicle
XX preparation from Gram-negative bacterial strain for vaccination of
XX microbial infections -
XX
XX Claim 46; Page 95; 128pp; English.
XX
XX The invention relates to a genetically-engineered outer membrane vesicle
XX (bleb) preparation from a Gram-negative bacterium for use as a vaccine.
XX The blebs of the invention are improved with respect to their
XX immunogenicity and toxicity by the introduction of one or more genetic
XX changes to the chromosome of the bacterium from which the blebs are
XX derived. The changes made include the upregulation of protective antigen
XX expression, the downregulation of immunodominant non-protective antigen
XX expression, and genetic changes which result in detoxification of the
XX Lipid A moiety of lipopolysaccharide (LPS). The invention also
XX encompasses modified Gram-negative bacterial strains from which the bleb
XX preparations are made, a vector suitable for performing recombination
XX events (for the generation of the modified bacterial strains),
XX bacterially-derived nucleic acid sequences used in such a vector, and an
XX immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole
XX cell vaccine suitable for paediatric use. The bleb preparation is useful
XX in the manufacture of a medicament for immunising a human host against a


```

4825 GGTAAAGCGTCTTACTCAACGGTATTATTTCCGATGTGAAGATTTCGG 4874
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
86  isGlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAla 102
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4875 TTTCCGGCAGCGCCCTGGTGTACGGTCAACGAGAACCAATTGGTCAAC 4924
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
103 GluIleAsnPheGluGlyAsnArgLeuIleProLeuGluGlyLeuGlnG1 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4925 CAGGTGCTGTTCACCGCAACCGCAAGATCAAGATGACAAAGCTGACGAC 4974
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
119 uGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaT 136
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4975 TGTGTGTCAGACGACGACCTTGGT .....CCTTATAGCCAGGAAC 5015
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGly 152
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5016 TCATCAGCGCCGATATCACCGCATCAACGACGACCTATCAGCCATTGGC 5065
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 TyrTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnAr 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5066 CGTAGTGAAGTTGAAGTGACGACGACCGCGCGTGGCCCTGGTCG 5115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValVala 186
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5116 CGTCAATCTCGTTTCGTCATCAATGAAGTGATCGACGAAATCGGCG 5165
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 spIleAsnIleIleGlyAsnGlnHisPheSerAspAlaLeuIleAsp 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5166 CTATCAACTTCGTCGTAATCATGCTTATGGCGATAGCCGCTTGGCGGT 5215
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
203 ValLeuAlaIleLysAspAsnLysIleAsnProLeuSer.....Ly 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5216 GTTATCAGCACCAAG .....AAGTCAATCCTTCTGCTTCGTGACCCG 5259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
216 sAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsnLeuA 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5260 TAAGGATGCTATAAGCAGCAGCAAGTTGAAGCTGACGAGAGCTCTGC 5309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
233 rGAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIleLysAspAla 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5310 GTCAGTCTATTACATCATGTTATGCGGATTTCCGGATTCAGCTTCG 5359
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
250 LysLeuAsnIleAsnGluAspLysAsnArgIlePheValIleSerLe 266
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5360 GATGCTCCTGAACGAGCAGACCAACGATACACCGTCAACATCACGCT 5409
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
266 uHisGluGlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnL 283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5410 TGATGAAGTCAAGCTGTACAAAGTTTCCGATATTATATGTCGAAAGCTCG 5459
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
283 euThr...TyrThrGlnAlaGluLeuAlaLeuLeuLysPheLysAla 298
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5460 TTGAAGCGTGGACCGCTAGTGTGAAGGGTCTGTGTGACGACATCGCCG 5509
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
299 GluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIleSe 315
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5510 GCGCGGCTCTACAGCGCCGCGAATCCAGAAATCGATGGAAGCCATCCA 5559
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
315 rThrLysPheGlyAspGlyTyrTyrTyrAlaGlnIleArgProValT 332
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5560 GCAGCGGTTTCGCCAAGGGTTATCCCTTTCGCCGCGTTCGTCGCGCG 5609
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
332 hrArgIleAsnAspGluSerArgThrValAspValGluTyrIleAsp 348
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5610 GCAACCGCATATGGCAATGGCACCATCGGCGTACCTATATGTTGAC 5659
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
349 ProValHisProValTyrValArgGlyIleAsnPheThrGlyAsnPhely 365
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
5660 CAGGTGACGCGCCCTATGTGACGCTATCGAAGTCAAGGCAATACAAG 5709
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
365 sThrGlnAspGluValLeuArgArgGlu 374
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

5710 GACACGCGATTACGTTATTTCGTCGCAA 5737
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:AAC81914
seq_documentation_block:
ID AAC81914 standard; DNA; 273254 BP.
XX AAC81914;
XX 27-FEB-2001 (first entry)
XX Chlamydia pneumoniae genome DNA.
KW Genome; diagnosis; vaccine; ds.
XX Chlamydia pneumoniae.
XX WO200027994-A2.
XX 18-MAY-2000.
XX 12-NOV-1999; 99WO-US26923.
XX 12-NOV-1998; 98US-0108279.
XX 08-APR-1999; 99US-0128606.
XX (REGC ) UNIV CALIFORNIA.
XX Stephens R, Mitchell W, Kalman S, Davis R;
XX WPI; 2000-376516/32.
XX Isolated nucleic acid for use in diagnostic and analytical methods
XX encodes genomic sequence of Chlamydia pneumoniae -
PS Claim 2; Page 128-320; 320pp; English.
CC This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC pneumoniae protein (p1), given in the specification. The isolated nucleic
CC acid is useful for diagnostic and analytical methods, such as,
CC hybridization-based assays or amplification-based assays. The protein may
CC be used for diagnostic purposes, for their enzymatic or structural
CC activity, or as a vaccine. The invention also describes (1) a probe
CC comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC cassette comprising N1 under the transcriptional regulation of a
CC transcriptional initiation region functional in an expression host, and a
CC transcriptional termination region; (4) a cell comprising an expression
CC cassette of (3) as part of an extrachromosomal element or integrated into
CC the genome of a host cell as a result of induction of the expression
CC cassette into the host cell, and the cellular progeny of the host cell;
CC (5) a method for producing a p1 comprising growing a cell of (4) where
CC the protein is expressed and isolating the protein free of other
CC proteins; (6) a purified polypeptide composition comprising at least 50
CC weight % of p1; and (7) a monoclonal antibody binding specifically to the
CC peptide of (6).
XX Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 other;

```

```

alignment_scores:
Quality: 395.00 Length: 881
Ratio: 0.855 Gaps: 36
Percent Similarity: 52.440 Percent Identity: 22.020
alignment_block:
US-09-701-711-2 x AAC81914/rev ..
Align seg 1/1 to reverse of: AAC81914 from: 1 to: 273254
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetal 17
||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
268517 ATCGGAATAAAGTTATCTTGCAATATCTATTCTAGCGTTAATCCAAC 268468

```

[illegible]

```

266766 AGGTGTC.....:
585  lYtPAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMet 601
: : |||||:||||:|||| ||| ||||| |||:
266758 ACTTGAATAGATTCTGTAGATAGTCTAGAACTCCAACTACAGGGATT 266709
: : |||||:||||:|||| ||| ||||| |||:
602  SerHisSerValAspLeuThrVal...GlyPheGlyAspLysThrHis.. 616
: : |||||:||||:|||| ||| ||||| |||:
266708 CGCGGGGGGTGACTTTTGTAGGTTTCTGGTTGGAGGAACCTATCATTT 266659
: : |||||:||||:|||| ||| ||||| |||:
617  .GlnLysValValTyrGlnGlnAsnIleTyrArgProPheIleLysLys 633
: : |||||:||||:|||| ||| ||||| |||:
266658 TACAAAACCTCTTTAAACAGCTCTATCTATAGAAAACCTTACGGCTAAAG 266609
: : |||||:||||:|||| ||| ||||| |||:
633  erValLeu.....ArgGlyTyrAlaLysLeu.....GlyTyrGlyAsn 645
: : |||||:||||:|||| ||| ||||| |||:
266608 GTATTTGAAATCAAGGGGAAGCTCAATTTATTAACCCCTATAGCAAT 266559
: : |||||:||||:|||| ||| ||||| |||:
646  .....AsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTy 658
: : |||||:||||:|||| ||| ||||| |||:
266558 ACTACAGCTGAAGGAGTTCCTGTGAGTGAGCGCTTCTTCCTAGGTGGAGA 266509
: : |||||:||||:|||| ||| ||||| |||:
658  rGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnA 675
: : |||||:||||:|||| ||| ||||| |||:
266508 GACTACAGATTGGGGATATAAATCTTTATATCGTCCAAAATACTCTG 266459
: : |||||:||||:|||| ||| ||||| |||:
675  laTyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluValVal 691
: : |||||:||||:|||| ||| ||||| |||:
266458 CT.....ACAGAACCCTCAG 266445
: : |||||:||||:|||| ||| ||||| |||:
692  GlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuPr 708
: : |||||:||||:|||| ||| ||||| |||:
266444 GGAGGACTCTCTTCGCTCCTTTATTCAGAGAGTTTCAATACCCCTC.. 266397
: : |||||:||||:|||| ||| ||||| |||:
708  oPheLysGlyAspTrpIleAspGln.....ValArgProValIlePheI 723
: : |||||:||||:|||| ||| ||||| |||:
266396 .....ATCAGACAACCTAATATAGTGCCTTTGTATTCT 266363
: : |||||:||||:|||| ||| ||||| |||:
723  leGlyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIle 739
: : |||||:||||:|||| ||| ||||| |||:
266362 TAGACTCAGGTTTGTG.....GGTTACAGAGATATAAGATT 266325
: : |||||:||||:|||| ||| ||||| |||:
740  AspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLy 756
: : |||||:||||:|||| ||| ||||| |||:
266324 TCGTTA..... 266319
: : |||||:||||:|||| ||| ||||| |||:
756  sAlaAlaAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerA 773
: : |||||:||||:|||| ||| ||||| |||:
266318 .....AAGATCTAGTAGTAGTG 266300
: : |||||:||||:|||| ||| ||||| |||:
773  laGlyValGlyAlaThrTriptyrThrProIleGly...ProLeuSerIle 788
: : |||||:||||:|||| ||| ||||| |||:
266299 CTGATTTGGTCTCGGCTTCGATGTAATGTAATGTTCTCGTTATGTTA 266250
: : |||||:||||:|||| ||| ||||| |||:
789  SerTyrAlaLysPro.....LeuAsnLysLysGlnAs 799
: : |||||:||||:|||| ||| ||||| |||:
266249 GGATTTGGTGGCCCTCCGTCACACCGAGACTTTGAATGGAGAAAAAT 266200
: : |||||:||||:|||| ||| ||||| |||:
799  nAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
: : |||||:||||:|||| ||| ||||| |||:
266199 TGATGTATCTCAGCGATTCTTCTTTGTTTGGGGGCGATGTC 266157
: : |||||:||||:|||| ||| ||||| |||:
seq_name: /SDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAx91990
seq_documentation_block:
ID AAX91990 standard; DNA; 1230025 BP.
XX
AC
XX
XX
DT 13-SEP-1999 (first entry)
XX
DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.

```

```

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; ss.
XX Chlamydia pneumoniae.
OS
XX WO9927105-A2.
XX PD 03-JUN-1999.
XX PF 20-NOV-1998; 98WO-IB01890.
XX PR 04-NOV-1998; 98US-0107078.
XX PR 21-NOV-1997; 97FR-0014673.
XX (GEST ) GENSET.
XX Griffais R;
XX WPI; 1999-357842/30.
XX Genome sequence of Chlamydia pneumoniae
XX Claim 1; Page 291-611; 1912pp; English.
XX The present sequence represents the complete genome of Chlamydia
CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
CC respiratory disease such as pneumonia and bronchitis and is thought
CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
CC encoded by the open reading frames of the C. pneumoniae genome (see
CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
CC Vectors containing C. pneumoniae nucleotide sequences can also be
CC used as immunogenic compositions, especially where the vector directs
CC the expression of a neutralising epitope of C. pneumoniae.
XX
XX Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
SQ

```

```

alignment_scores:
  Quality: 395.00      Length: 881
  Ratio: 0.855        Gaps: 36
  Percent Similarity: 52.440  Percent Identity: 22.020
alignment_block:
US-09-701-711-2 x AAX91990 ..
Align seg 1/1 to: AAX91990 from: 1 to: 1230025
1 MetArgAsnSerTyrPheLysGlyPheGlnValSerAlaMetThrMetal 17
|||||:||||:|||| ||| ||||| |||:
348099 ATGGCAATAAAGTTATCTTCCAAATATCTATTCTAGCGTTAATCCAAC 348148
: : |||||:||||:|||| ||| ||||| |||:
348149 CCCTTAACTTTATTTCTACTGAAAAGTTAAAGAGGCCATGTGGTGG 348198
: : |||||:||||:|||| ||| ||||| |||:
33 laAsnAspIleThr...IleThrGlyLeuGlnArgValThrIleGluSer 48
: : |||||:||||:|||| ||| ||||| |||:
348199 TAGACTCTATCACAAATCAATACGGAAGGAGAAAATGCTTCAATAAACAT 348248
: : |||||:||||:|||| ||| ||||| |||:
49 LeuGlnSerValLeuProPheArgLeuGlyGlnValValSerGluAsnG1 65
: : |||||:||||:|||| ||| ||||| |||:
348249 CCCTTACCAATTAAGACCAGAGTGGGCTCTTTTCTCAATTAGA 348298
: : |||||:||||:|||| ||| ||||| |||:
65 nLeuAlaAspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspV 82
: : |||||:||||:|||| ||| ||||| |||:
348299 TTTTGTATGACAGCTTGAGAATTTCTAGCTAAAGAAATACGACTCTGTTGAGC 348348
: : |||||:||||:|||| ||| ||||| |||:
82 alGlnValTyrHisGlnGluGlyArg.....IleIleTyrGlnValThr 96
: : |||||:||||:|||| ||| ||||| |||:
348349 CTAAGTAGAATTTTCTGAGGGAAGAACTAATACATAGCCCTTCACCTAATA 348398
: : |||||:||||:|||| ||| ||||| |||:

```


350058 ACTACAGCTGAAGAGTTCCTGTCAGTGAGCGCTTCTCTAGTGGAGA 350107
658 rGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnA 675
350108 GACTACAGTTCGGGGATATAAATCCCTTTATTATPCGGTCCAAATACTCTG 350157
675 laTyrLeuThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluValVal 691
350158 CT.....ACAGAACCTCAG 350171
692 GlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuLeuProLeuPr 708
350172 GGAGGACTCTCTCGCTCCTATTATTCAGAAGAGTTTCAATACCTCTC.. 350219
708 oPheLysGlyAspTyrPheAspGln.....ValArgProValIlePheI 723
350220ATCAGAACCTATATTTAGTCGCTTTGTATCT 350253
723 leGluGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIle 739
350254 TAGACTCAGGTTTGTG.....GGTTACAAGAGTATAAGATT 350291
740 AspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAla 756
350292 TCCTTA..... 350297
756 saAlaAlaAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSera 773
350298AAGATCTCAGCTAGTAGTG 350316
773 laGlyValGlyAlaThrTrpTyrThrProIleGly...ProLeuSerIle 788
350317 CTGGATTGTGCTCGCGCTTCGATGTAATGAATAATGTTCTCTGTATGTTA 350366
789 SerTyrAlaLysPro.....LeuAsnLysLysGlnAs 799
350367 GGATTGTGTGGCCCTCGCTCCAACCGAGACTTTGAATGGAGAAAAAT 350416
799 nAspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
350417 TGATGATCTCAGCGATTCTCTTCTGCTTGGGGGCGATGTC 350459
seq_name: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.AAS86808
seq_documentation_block:
ID AAS86808 standard; cDNA; 3051 BP.
XX AAS86808;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #22612.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.

DR P-PSDB; ABG22621.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 1; SEQ ID No 22612; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3051 BP; 710 A; 799 C; 820 G; 722 T; 0 other;

alignment_scores:
Quality: 393.50 Length: 460
Ratio: 1.555 Gaps: 14
Percent Similarity: 55.000 Percent Identity: 28.696
alignment_block:
US-09-701-711-2 x AAS86808 ..
Align seg 1/1 to: AAS86808 from: 1 to: 3051
211 IleAsnProLeuSerLysAlaAspArgTyrThrGln..... 222
||| |||:||||| :|||:|||||: 1795 ATCGCCCCATCAGTAACACACAGGACACACACATCCA 1844
223GluL 224
1845 GGAAGTGCCAGAACAGGCTCAGGCACATGATGCGGATTCATCCCGATA 1894
224 ysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGly 240
|| |||:||||| :|||:|||||: 1895 AATCCGCGCAACCAACAGGTCAACGCCAGGAGGAGATAACCTGGCT 1944
241 PheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLy 257
||||| ||| |||:||||| :|||:|||||: 1945 TTTGTTGTTTTCATATGCGGATAGCCGCATCGCTAGGTGATGGAGC 1994
257 sAsnArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgp 274
1995 T.....GAACAACAGCAAGAAAGTT 2014
274 heGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeu 290
|||:||||| :|||:|||||: 2015 TCAACCAG.....AACGAACGGCAGTTT 2037
291 GluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLe 307
||| :|||:||||| :|||:|||||: 2038 GAAATGTCTTACCTCTCGGCGCGCTGCGTCCCGTTCAC..... 2079
307 uGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspGlyTyrT 324

```
2080 .....CAGAACGGCATAGGTAGCAAA..... 2100
324 yTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThr 340
2101 .....CAGATAGAACAGAAAT 2118
341 ValAspValGluTyrTyrIleAspPro.....ArgIleAsnPhe. 360
2119 GCA...GTGGCTCATCAGGTAGATCCAAATCCGAAGATTTTGGTTCCGC 2165
350 .....ValHisProValTyrValArg..... 360
2166 CTGCAATCGTGGTGGCGGTTCGTGGCGGTGGCGAGTCGCGTACCATACC 2215
361 ThrGlyAsnPheLysThrGlnAspGluValLeuArgArgGluMetArgG1 377
2216 TCG.....AAAGATCCGTCCTGCGTCGCGGAATCGGTCA 2250
377 nLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaA 394
2251 GATGGAAGGTGATG..GCTGGGAGCGATCTGCTGCATCAGGTAAGGAGC 2299
394 rGLeuMetArgThrGlyPhePheLysHisValThrValAspThrArgPro 410
2300 GTCTGAATCGTCTGGCTTCTTTGAACATGTCGTACCGATACCCAAAGT 2349
411 ValProAsnSerProAspGlnValAsnValAsnPheValValGluGluG1 427
2350 GTTCCGGGTAGCCCGACCGACCGTTCGTCTACAGGTAAAGAGCG 2399
427 nProSerGlySerThrIleAlaAlaGlyTyrSerGlnSerGlyGlyV 444
2400 CAACACCGGTAGCTCAACTTGTGTTGTTACGGTACTGAAAGTGGCG 2449
444 alThrPheGlnPheAspValSerGlnAsnAsnPheMetClyThrGlyLys 460
2450 TGAGCTTCAGGCTGGTGTGACGAGGATACTGTTAGGTACAGGTTAT 2499
461 HisValAsnAlaSerPheSerArgSerGluThrArgGluValTyrSerLe 477
2500 GCTGTTGGTATCAACGGGACCAAAACGATTACACAGACCTATGCTGAAC 2549
477 uGlyMetThrAsnProTyrPheThrValAsnGlyValSerGlnSerLeu 494
2550 CTCGGTAACCAACCGGTACTTTCACCGTAGATGGCGTAAAGCCTCGGTC 2599
494 erGlyTyrTyrArgLysThrLysTyrAspAsnLysAsnLysSerAsnTyr 510
2600 GTCTCTCTTAATGACTTCACGCGAGATGACCGCGACCTGTCGCACTAT 2649
511 ValLeuAspSerTyrGlyGlySerTyrGlyTyrProIleAspG1 527
2650 ACCAACAAGAGTTATGTTACAGACGTGACGTGGCTTCCCGATTACGA 2699
527 uAsnGlnArgIleSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisG 544
2700 ATATAACCTCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 2745
544 LyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAla..... 558
2746 .....TCCACATGCGACCTCAGGTTCGGATGGG 2775
559 .....AspGlyLysIleGlnValAspAsnAsnGlyIlePr 571
2776 CGTATCTGTACTCTATGGTGAACATCCGAGCACCTCTGTACGAGTAA 2825
571 oAspPheLysHisAspTyrThrTyrAsnAlaIleLeuGlyTrpAsnT 588
2826 CAGCTTAAAAACGACGACTTCACGTTCAAC.....TATGTTGGACCT 2869
588 yrSerSerLeuAspArgProValPheProThrGlnGly.....MetSer 602
|||||
```

```
2870 ATAACAAGCTTGACCGTGGTTACTTCCCGACAGATGGTTACCGTGTCAAC 2919
603 HisSerValAspLeuThrValGlyPheGlyAspLysThrHisGlnLysVa 619
2920 CTGACCGGTAAGTGAACATTCCTGGATCGGATACCAATACATACTACAAAGT 2969
619 lValTyrGlnGlyAsnIleTyrArgPro 628
2970 GACGTTAGACACGGCGACTTATGTGCG 2997
seq_name: /SIS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.AAZ61502
seq_documentation_block:
ID AAZ61502 standard; DNA; 2550 BP.
XX
AC AAZ61502;
XX
DT 19-JUN-2000 (first entry)
XX
DE DNA encoding the CPN100111 polypeptide.
XX
KW CPN100111; Chlamydia infection; immune response; vaccine; ss.
XX
OS Chlamydia pneumoniae.
XX
FH Key Location/Qualifiers
FT CDS 101..2464
FT sig_peptide 101..352
FT sig_peptide 101..352
XX
PN WO200011183-A2.
XX
PD 02-MAR-2000.
XX
PF 18-AUG-1999; 99WO-IB01449.
XX
PR 20-AUG-1998; 98US-0097187.
PR 20-AUG-1998; 98US-0097188.
PR 20-AUG-1998; 98US-0097189.
PR 20-AUG-1998; 98US-0097190.
PR 20-AUG-1998; 98US-0097195.
PR 20-AUG-1998; 98US-0097196.
PR 20-AUG-1998; 98US-0097197.
PR 27-AUG-1998; 98US-0097191.
PR 17-AUG-1999; 99US-0376770.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP;
XX
DR WPI; 2000-224703/19.
DR P-PSDB; AAY69362.
XX
PT Novel antigens and corresponding DNA molecules that can be used to
PT prevent, treat and diagnose disease caused by Chlamydia infection in
PT mammals, especially humans -
XX
PS Claim 1; Fig 1A-F; 201pp; English.
XX
CC AAZ61502-09 encode Chlamydia pneumoniae polypeptides. The polypeptides
CC are present in the bacterial membrane structure, in the external
CC vicinity of the membrane structure, in the inclusion membrane
CC structure, in the external vicinity of the inclusion membrane structure,
CC and in the cytoplasm of the infected cell. The polypeptides may be
CC used to prevent, treat and detect the presence of Chlamydia infection
CC and/or the presence of Chlamydia in a sample. The polypeptides may
CC also be used to induce an immune response in a mammal. The vaccine
CC vector comprising the polynucleotides is used to induce an immune
CC response in a mammal. Antibodies directed against the polypeptides
CC may also be used therapeutically to treat and/or prevent a Chlamydia
CC infection.
XX
```

[illegible]

```

1667 AACAGAGCATATCTAAGATTATGCTCCAAACCTATGGCGGAAGT 1716
519 userTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuA 536
1717 CACACACAGTATATCTTGAACGACACCTGAATACGGTCTATTTAT 1766
536 snAlaAspAsnThrLysLeuHisGlyGlyArg...PheMetGlyIleSer 551
1767 GAGGAAGTCAACAGAGTTTACATGAAACGTAAGTTCTCTCTAGGGCCA 1816
552 AsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnAs 568
1817 AATATA.....GACAGCAATAAAGGATTGTCTCTGCTGC 1851
568 nGlyIleProAspPheLysHisAspTyrThrTyrAsnAlaIleLeuG 585
1852 AGGTGTC.....A 1859
585 lyTrpAsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMet 601
1860 ACTTGAATTACGATTCTGTAGATAGTCTTGAACCTCAACTCAGGGATT 1909
602 SerHisSerValAspLeuThrVal...GlyPheGlyAspLysThrHis.. 616
1910 CGCGGGGGGCTGACTTTTGGAGTTTCTGTTGGAGGAACCTATCATTT 1959
617 GlnLysValValTyrGlnGlnAsnIleTyrArgProPheIleLysLys 633
1960 TCAAAACACTCTTTAAACAGCTCTATATAGAACTATAGAACTACGCGTAAAG 2009
633 erValLeu.....ArgGlyTyrAlaLysLeu.....GlyTyrGlyAsn 645
2010 GTATTTGAAATCAAGGGAAGCTCAATTATTAAACCTATAGCAAT 2059
646 AsnLeu.....ProPheTyrGluAsnPheTyrAlaGlyGlyTy 658
2060 ACTACAGCTGAAGGAGATCTCTGAGTGAAGCGCTTCTCTAGGTGAGA 2109
658 rGlySerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnA 675
2110 GACTACAGTCGGGATATAAATCTTTTATTCGTGTCCTCAAAATACTCG 2159
675 laTyrLeuThrAlaArgArgGlyGlnGlnThrLeuGlyGluValVal 691
2160 CT.....ACAGAACCTCAG 2173
692 GlyGlyAsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuPr 708
2174 GGAGGACTCTTCGCTCTTATTTCAGAGAGTTTCAATACCCCTC... 2221
708 oPheLysGlyAspTrpIleAspGln.....ValArgProValIlePheI 723
2222 .....ATCAGACAACCTAATATTAGTCTTGTATCT 2255
723 leGluGlyGlyGlnValPheAspThrThrGlyMetAspLysGlnThrIle 739
2256 TAGACTCAGGTTTCTC.....GTTTACAAGAGTATAAGATT 2293
740 AspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaIy 756
2294 TCGTTA..... 2299
756 sAlaAlaAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerA 773
2300 .....AAGACTCTACGTAGTAGTG 2318
773 laGlyValGlyAlaThrTrpTyrThrProIleGly...ProLeuSerIle 788
2319 CTGGATTGGTCGCGCTCGATGTAATGAATATGTTCTGTTATGTTA 2368
789 SerTyrAlaLysPro.....LeuAsnLysLysGlnAs 799

```

```

2369 GGATTTGGTGGCCCTTCGTCACACGAGACTTTGAATGGAGAAAAAT 2418
799 naspGlnThrAspThrValGlnPheGlnIleGlySerValPhe 813
2419 TGATGATCTACGAGTATCTTCTTTAGGGGCGCATGTC 2461
seq_name: /SID51/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAZ01425
seq_documentation_block:
ID AAZ01425 standard; DNA; 1038602 BP.
XX
AC AAZ01425;
XX
DT 07-OCT-1999 (first entry)
XX
DE Complete genome sequence of Chlamydia trachomatis.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
KW nongonococcal urethritis; epidymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
XX
OS Chlamydia trachomatis.
XX
FN WO9928475-A2.
XX
PD 10-JUN-1999.
XX
PF 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST ) GENSET.
XX
PI Griffais R;
XX
DR WPI; 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
PS Claim 1; Page 373-656; 1755pp; English.
XX
The present sequence represents the complete genome of Chlamydia
trachomatis. Open reading frames (ORFs) of the genome encode
polypeptides AY36754-Y37949. The polypeptides can be used as vaccines
against Chlamydia trachomatis. Antisense and ribozyme sequences can also
be used to control growth of the microorganism. Chlamydia trachomatis is
responsible for a large number of diseases, e.g. eye diseases such as
conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
conjunctivitis; genital diseases such as nongonococcal urethritis,
CC epidymitis, cervicitis, salpingitis, perihhepatitis, bartholinitis;
CC pneumopathy in breast feeding infants; and venereal
CC lymphogranulomatosis. The polypeptides of the invention may be of use in
treating these diseases.
XX
SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

alignment_scores:
Quality: 386.50 Length: 907
Ratio: 0.880 Gaps: 35
Percent Similarity: 48.401 Percent Identity: 21.058

alignment_block:
US-09-701-711-2 x AAZ01425/rev ..
Align seg 1/1 to reverse of: AAZ01425 from: 1 to: 1038602
14 MetThrMetAlaValMetMetValMetSerThrHisAlaGlnAlaAlaAs 30
206407 CTGCNACTCGCTGTTTACTGTGCTCACC..... 206378

```

30 pPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleG 47
||| : : : : : ||| : : : : :
206377 .TTTTCACAAAGTCTTCTGTTCAACTTCAGAGGACGTATGGTCGTAG 206329
47 luSerLeu.....GlnSerVal 52
||| : : : : : ||| : : : : :
206328 AGCTATACCACTTACGACTCAAGGAGAGAACTCAAAATAAACGAGCT 206279
53 LeuPro.....PheArgLeuGlyGlnValSerGluAsnGlnLe 66
: : : : : ||| : : : : : ||| : : : : :
206278 ATTCTTAATAAACAAGCAAGGACGTTGTTCTCTCAAGCAGATTT 206229
66 uAlaAspGlyValLysAlaLeuThrGlyAsnPheSerAspValG 83
: : : : : ||| : : : : : ||| : : : : :
206228 TGATGAAGATCTAAGAACACTT.....TCGAAAAATTTGATCGAGTAG 206185
83 ln.....ValTyrHisGlnGluGlyArgIleIleTyrGln.....Val 95
: : : : : ||| : : : : : ||| : : : : :
206184 AGCCTATCGTAGAGTTTCGTAATGGACAAGCTGTGATCTCTCTGATTCTG 206135
96 ThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuI 112
||| : : : : : ||| : : : : : ||| : : : : :
206134 ACGCAAAATCTGTTATCAGAGAGATCAATATTTTCAGAAATGAAGCTAT 206085
112 eProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValG 129
||| : : : : : ||| : : : : : ||| : : : : :
206084 CCCCACTCAATAAATT..... 206069
129 lyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThr 145
||| : : : : : ||| : : : : : ||| : : : : :
206068CTGAAA.....ACTTTAGAGCTTTATAAAATGATCTTTTT 206033
146 AsnGln.....TyrIle 149
: : : : : ||| : : : : : ||| : : : : :
206032 GATCGGGAATTTCTTTAAATTTTGATCGCTAAGAACTCTTTATTT 205983
149 eSerGlnGlyTyrTyrAsnThrGluIleThrVal..... 160
: : : : : ||| : : : : : ||| : : : : :
205982 GAAACGAGGCTACTACGATCTCAACTCTCTCTATTCATAATCATAATG 205933
161 ..LysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMetThrPhe 176
||| : : : : : ||| : : : : : ||| : : : : :
205932 AGAAAGAGGCTTTATCGAT.....ATTCCATCGAGATT 205898
177 AlaGluGlyLysProAlaArgValValAspIleAsnIleIleGlyAsnG 193
||| : : : : : ||| : : : : : ||| : : : : :
205897 AAGAAGAGCGTCACGCTCCGATAAAAAAATTAACGATTTCCGGAATAC 205848
193 nHisPheSerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsnL 210
: : : : : ||| : : : : : ||| : : : : :
205847 GCGAACAGAGCATCAGACTTAGTGACATGTTTAACTAAACAATACT 205798
210 ysIleAsnPro.....LeuSerLysAlaAspArgTyrThrGlnGluLys 224
: : : : : ||| : : : : : ||| : : : : :
205797 CCACACACAGAGCTGGTTTCACTGGTCCGAGGTGATCATCCGAGCATG 205748
225 LeuValThrSerLeuGluAsnLeuArgAlaIlyTyrLeuAsnAlaGlyPh 241
: : : : : ||| : : : : : ||| : : : : :
205747 GTAGACAAGACTTTTGTGCTATCACAATACTTCCAAATAAAGGATA 205698
241 eValArgPheGluIleLysAspAlaLysLeuAsn.....IleAsnGluA 256
: : : : : ||| : : : : : ||| : : : : :
205697 TGCT.....GATGCTAAGTAAGCAAGAGGCTCTCTACAG 205663
256 sp...LysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
||| : : : : : ||| : : : : : ||| : : : : :
205662 ATGCTAAAGGAAACATTACTTTGCTTATCGTTGTAGACAAAGGACCTTTA 205613
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
||| : : : : : ||| : : : : : ||| : : : : :
205612 TACACATTAGGTCACGTACATATAGAGGATTTCACAGCGTTATCCAAAG 205563
288 aGluLeuGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnA 305
||| : : : : : ||| : : : : : ||| : : : : :
205562 ACTCTCGATAAACAACACTATTGGTTGGACCTAACTCTTATATGCCACG 205513
305 laMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAsp 321
: : : : : ||| : : : : : ||| : : : : :
205512 ATAAAAATTTGGACTGGAGCACAAAAGATTCTAGCGCATACGCTAGATAT 205463
322 GlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspGluSe 338
||| : : : : : ||| : : : : : ||| : : : : :
205462 GGCTACGTG.....AA 205452
338 rArgThrValAspValGluTyrTyrIleAspProValHisProValTyr. 354
: : : : : ||| : : : : : ||| : : : : :
205451 CACTAACGTTGATGTCCTCTCAGCGCACCCCACTCTACCTGTTTACG 205402
355ValArgArgIle 358
: : : : : ||| : : : : : ||| : : : : :
205401 ATGTTACTATCGAGTGAGTGAGGATCTCCCTACAAAATCGGGTTAATT 205352
359 AsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArgGluMe 375
: : : : : ||| : : : : : ||| : : : : :
205351 AAAATCAAGGAACACTCATCTAAGCATGATGTGATTTGCATGAGAC 205302
375 tArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerA 392
: : : : : ||| : : : : : ||| : : : : :
205301 TAGCTTTTCCCTCGAGACACTTTTATAGATTAAAACTAGAGATACAG 205252
392 rgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAspThr 408
: : : : : ||| : : : : : ||| : : : : :
205251 AGACTCGTTTACGCAACACCGGCTACTTTAAAGTGTAAGTGTCTATACG 205202
409ArgProValProAsnSerProAspGlnValAspVa 420
||| : : : : : ||| : : : : : ||| : : : : :
205201 GTTCGTTCCCAATTAGATCCTCTCGATTCTAACGCACCTTTATCGAGATGT 205152
420 laAsnPheValValGluGluGlnProSerGlySerSerThrIleAlaAlaG 437
: : : : : ||| : : : : : ||| : : : : :
205151 TTTTATGAAGTCAAAGAGACTGAAACAGGAAATCTTGGGCTATTCTTAG 205102
437 lyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsn 453
||| : : : : : ||| : : : : : ||| : : : : :
205101 GATTCACTCCATTGACCATTTATTGGAGAGCAGAAATTCAGAAAAGC 205052
454 AsnPhe..... 455
||| : : : : : ||| : : : : : ||| : : : : :
205051 AACTTTGATTATTGGAGCCCGAACTTTCTCAAAAAGGATTCAAATC 205002
456MetGlyThrGlyLysHisValAsnAlaSerPheSerArgSerGluT 471
||| : : : : : ||| : : : : : ||| : : : : :
205001 TTTAAGAGGTGGTGGAGAATACCTCTTCTCTAAAGCTAATTTAGGAGATA 204952
471 hrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheThr..... 485
||| : : : : : ||| : : : : : ||| : : : : :
204951 AGGTCACCGATTACACTGTTAAATGGAGAACCAACCACTCTCTTAAATACC 204902
486ValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysTh 500
: : : : : ||| : : : : : ||| : : : : :
204901 CCTTGGATTCTGGAGTAGAATTAGAT..... 204875
500 rLysTyrAspAsnLysAsnIleSer...AsnTyrValLeuAspSerTyrG 516
||| : : : : : ||| : : : : : ||| : : : : :
204874 .AAATCAATTAATAAGCTTTATCAAAAGACTACTCTGTGGATACCTATG 204826
516 lyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSer 532
||| : : : : : ||| : : : : : ||| : : : : :
204825 GAGGAATATCAGTACCACCTACATTTCTT.....AACGATAAGTAAAA 204782
533 PheGlyLeu.....AsnAlaAspAsnThrLysLeuHisGlyGlyArgPh 547
: : : : : ||| : : : : : ||| : : : : :
204781 TATGGGATGATTATACCGTGGTAGCCCAACAAAGCTTAAGTTTGGCGCAAAA 204732
547 eMetGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleG 564

69 CATAGGATCTCGAAA.....GGACAGTCTCTCGGATCTGACCCCTCAGG 112
 35 spile.....
 113 ATACTATATATAGCCCTGAAATATCTATGCCAAGCCTATTCTATAGACC 162
 38 lIeThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuPr 54
 163 ATAGCATCTATTGAG.....ATCGAGGGAATGAGGCTCTTCGATGA 203
 54 oPheArgLeu.....GlyGlnValValSerGluAsnGlnLeuA 67
 204 CTTTGTCTTGGCAATCTTTCAGCGTTGGCTGATGATGAAGTCTCTGA 253
 67 laAspGly.....ValLysAlaLeuTyrAlaThrGly 77
 254 TTCCTGGAGATCCCATGCTGCTGCGGTGAATAGAAATATATGCTCAGGCG 303
 78 AsnPheSerAspValGlnValTyrHisGlnGlu.....GlyArgIleI 92
 304 TACTTCTCAATGTGCGAATCATCGCGGATAAATATATGTCGCAATAAAGT 353
 92 eTyrGln.....ValThrGluArgProLeuIleAlaGluIleAsnP 106
 354 CTATCTGAAATCATGTGTCAGTGAACGTCCTCGCATCAGTAAGTTACTT 403
 106 heGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLys 122
 404 TTAGCGGGGTAAGAAGTCTGAGAGAGAAGATCTTGAA.....ATG 444
 123 AsnAlaGlyLeuAlaValGlnProLeuLysGlnAlaThrValGlnMe 139
 445 AAAATCGGTCTTCGCGAGGGATTCAGATGACCAAGAAATAATGAAGACAA 494
 139 tIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnT 156
 495 GGTACGGCAAAATCGTACAGAAGATATTTAGTCAGAAAGGTTATCGCGAT 544
 156 hrGluIleThrValLysGln.....ThrMetLeuAspGlyAsnArgVal 170
 545 CCAGCATACGGATACGACGAGGACCGGATCTTCCAAAGATGGCTTGTGC 594
 171 LysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspI 187
 595 AATGTCTTATCTCGATTGAGAAGAAAACCAACCAAGGTGAATGAAT 644
 187 eAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVal 204
 645 TTATTTTTCGGCAACAAGGCCCTTAGCAATCATTAAGCTA.....AGAA 688
 204 euAlaIleLysAspAsnLysIleAsn..... 212
 689 TGGCGATGAAGAACACCAATGCAATTCAGTCTTAGAAGCATATTCCG 738
 213ProLeuSerLysAlaAspArgTyrThrGlnGluLysLe 225
 739 TCATCTTTCTTGAACCTTTTGTAGTACTATAAGTTTGTGGAAGAGAGCTA 788
 225 uValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheV 242
 789 CCGTGAAGATTTTGGCGGATTCATGAGAAGATATCAGGAATATGGATATC 838
 242 aArgPheGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsn 258
 839 GTGATGCTGAATATACTACCGACAGTGTGCTGAAGGCTCTCAGCGCAAA 888
 259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheG 275
 889 ACAGTGGATATTATCTCAACATCGAAGAGGGGCGAGAAGTATTATATAA 938
 275 yGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluA 292
 939 GGATGTCAACTTTTGGGCAATTCAACATATCCATCCGAGTAGTATTTGGAAC 988

292 laLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeuGlu 308
 989 GAGTGTCTCGGAATAAATCCGGAGATGTGTACAATCAGAGACGATTGGCT 1038
 309 GlnThrThrAsnAsn.....IleSerThrLysPheGlyAspAs 321
 1039 AAGCGTCTCAATGAAGATGAAGATGCTGTGGGGAACCTGTACTATAACAA 1088
 321 pGlyTyrTyrTyrAlaGlnIleArgProVal...ThrArgIleAsnAspG 337
 1089 TGGCTATATTTTTCGTTGGTGCATCCCGTGGAAACAATGTAGTGGGG 1138
 337 luSerArgThrValAspValGluTyrTyrIleAspProValHisProVal 353
 1139 ATTCTCTTTCGCTTGATATTCTG.....ATAGCGGAGGGGAAGCAGGCC 1182
 354 TyrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVa 370
 1183 AATATCAATAAGGTGATCATCAAGAAATATCTGCTGTACGAAGACGT 1232
 370 lLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnL 387
 1233 AGTACCCCGAGAGCTTTACACAAGCCCGCCAGCTCTTTAGTCGGGAGG 1282
 387 ysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHis 403
 1283 ATATCATTAATCTATCTCTCATCAATCAGCTTGGGCATTTTCGATGCC 1332
 404 ValThrValAspThrArgProValProAsn...SerProAspGlnValAs 419
 1333 GAAAAATCTATTCCTCGCTCCGATTCCTCAATCCGAAACAGGAACAGTGA 1382
 419 pValAsnPheValValGluGlnProSerGlySerSerThrIleAlaA 436
 1383 TATAGAGTATGATTGGTCCGCGTAGCAGTACCAATTCGGAGCTTTCG 1432
 436 laGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerGln 452
 1433 TCGGTGGAGTCAGTCC...GGACTCTGTTCGAGAGGCCATTAAAGTTC 1479
 453 AsnAsnPhe..... 455
 1480 AGCAACTTCTCTGTCGGCAACTTGCCTCCATCCCTCGATGTATAAGAAAG 1529
 456MetGlyThrGlyLysHisValAsnAlaSerPhe...SerA 468
 1530 GATCATTCGCAAGGGGATGGGCAACACATATCCTCAGTGCCTCAGACCA 1579
 468 rgSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPhe 484
 1580 ATGGAAGTACTATCAGCAGTAGTAGTGTACATTTATGGATCCATGGTT 1629
 485 ThrValAsnGlyValSerGln...SerLeuSerGlyTyrTyrArgLysTh 500
 1630 GGGGCAAGCGCGCGGATATGTTCAGCTTCAGTTCAGTTCATTATTCACAG 1679
 500 r.....LysTyrAspAsnLysAsnIleSerAsnTyrVal. 511
 1680 TAGCGCGATTGACTCCCAAGTCTACAAATAGCAATGCCGCAACTACTATA 1729
 511 511
 1730 ATGCCTACTATAATAGCTACTACAACAACTATAATAGTTATTACAACGGT 1779
 512 LeuAspSerTyrGlySerLeuSerTyrGlyTyrProIleAspGluAs 528
 1780 ATGTGCAACTATACCGCGCACCTC...TATACTCAGGCCAGCGATCCGGA 1826
 528 nGlnArgIleSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyG 545
 1827 TCGTTCGCTT..... 1836

```

545 lYArgPheMetGlyIleSerAsnValLysGlnLeuMetalAlaAspGlyGly 561
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
1837 ..CAGATGTTAGGTAAGTTCG.....ATCGTTACGGTAAG 1869
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
562 LysIleGlnValAspAsnGlnLysLeuProAspPheLysHisAspTyrTh 578
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
1870 CGTTGACTGCGCGGACAAATGGTTCAGATTATATCTTCTCTGAACATA 1919
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
578 rThrTyrAsnAlaIleLeuGlyTyrAsnTyr..... 588
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
1920 CACCTACTATAGATGCGCAATATGGAGTACAAATACCTTCCAAATATCC 1969
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
589 .....Ser 589
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
1970 ATCATGGCTCGCTAAATGATCTCAACTTGGAGCTCGCTCTCTCGTACT 2019
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
590 SerLeuAspArgProValPheProThrGlnGlyMetSerHisSerValAs 606
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2020 TCCATCGATAATCTTATTTACAGAAAGCGGATTCATGGTTTC 2069
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
606 pLeu..... 607
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2070 TGTGCTGCTACTCTCTCTTATCTTTTGGGACAATCATGACTATGCCA 2119
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
608 .....ThrValGlyPheGlyAspLys.....ThrHisGlnLys 618
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2120 GCCAGAACCTCAGCGTAAGCATGCTTACAGATTATCGATCATCACAAG 2169
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
619 ValValTyrGlnGlyAsnIleTyrArgProPheIleLys..... 631
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2170 TGGAGTTTAGAGGACGAGTTTACTCCATGCTCAATCTGTACGCCA 2219
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
632 .....LysSerValLeuArgGlyTyrAlaLysLeuG 642
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2220 TAAATATACACCGTGCCTCATGAGTCGAGTGGAGGACGACTTCTGGTT 2269
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
642 lyTyrGlyAsnAsn.....LeuProPheTyrGluAsnPheTyrAlaGly 656
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2270 CGTATAATTCCAATAAGAAATCTCTTTC...GGTACTTTCTATATGGGA 2316
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
657 Gly.....TyrGlySerValArgGlyTyrAspGlnse 667
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2317 GGTGATGATGTCCAGCTATTATGGT.....GGCTACATCAATGA 2357
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2358 GACTATAGGT.....TTGCGTGGTTATA 2380
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
684 lnThrThrLeuGlyGluValValGlyGlyAsn.....Ala 695
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2381 AGAAC.....GGATCTATTCGCGTAAATAACTACGACTATGCTATGCT 2424
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
696 LeuAlaThrPheGlySerGluLeuLeuProLeuProPheLysGlyAs 712
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2425 TATATGCGCTTACGATGGAAGTACGTTCCCGATTCTGTTGAAACTC 2474
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
712 p.....TrpIleAspGlnValArgProValIlePheIleGluGly 726
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2475 ATTCAATGCTGGCTC.....TTAGCTTTTCCCGAAGCAG 2509
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
726 lyGlnValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThr 742
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2510 GCAATGCGTGG.....CGCAGTATCGAC..... 2532
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
743 GlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaLaAs 759
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2533 .....AATTATATCCCTTTAA 2549
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
759 nArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValG 776
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2550 C.....CTGAGCGATCGCGCGGTGTAG 2572
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
776 lyAlaThrTrpTyrThrPro...IleGlyProLeuSerIleSerTyrAla 791
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||

```

```

2573 GATTGCGGTAAACGTTACGATGGTCCGAATGCTCGGTATCGATTGGGA 2622
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
792 .....LysProLeuAsnLysLysGlnAsnAspGlnThrAspTh 804
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2623 TATGCGTTTGACCGTCCGACAAATCTCTACAGCGAGGAGGAAGCAAT... 2670
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
804 rValGlnPheGlnIleGlySerValPhe 813
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||
2671 .GTCCACTTTGCTCGCAGCAGAGTTTC 2697
      ::::::::::: ||| ::::: ||| ::::: ||| ::::: |||

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.AAX91713
seq_documentation_block:
ID   AAX91713 standard; DNA; 2673 BP.
XX
AC   AAX91713;
XX
DT   25-AUG-1999 (first entry)
XX
DE   Porphyromonas gingivalis protein PG36 encoding DNA.
XX
KW   Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX   vaccine; antigenic; ds.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO9929870-A1.
XX
PD   17-JUN-1999.
XX
PF   10-DEC-1998; 98WO-AU01023.
XX
PR   04-AUG-1998; 98AU-0005028.
XX   10-DEC-1997; 97AU-0000839.
XX   31-DEC-1997; 97AU-0001182.
XX   30-JAN-1998; 98AU-0001546.
XX   10-MAR-1998; 98AU-0002264.
XX   09-APR-1998; 98AU-0002911.
XX   23-APR-1998; 98AU-0003128.
XX   05-MAY-1998; 98AU-0003338.
XX   22-MAY-1998; 98AU-0003654.
XX   29-JUL-1998; 98AU-0004917.
XX
PA   (CSLC-) CSL LTD.
XX
PI   Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
XX   Ross BC, Rothel LJ, Webb EA;
XX
DR   WPI; 1999-385613/32.
XX   P-PSDB; AAY34495.
XX
XX   Antigenic Porphyromonas gingivalis peptides for preventing
XX   gingivitis
XX
PS   Claim 12; Page 215-216; 588pp; English.
XX
CC   AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX   Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX   AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX   isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX   activity with a vaccine mechanism of action. The PG polypeptides can be
XX   used as vaccines especially against Porphyromonas gingivalis. Probes can
XX   be used to detect Porphyromonas gingivalis in standard hybridisation
XX   assays. Porphyromonas gingivalis is involved in periodontal disease
XX   especially gingivitis.
XX
SQ   Sequence 2673 BP; 747 A; 549 C; 638 G; 739 T; 0 other;

```

```

alignment_scores:
Quality: 378.50      Length: 973
Ratio: 0.795        Gaps: 43

```


Percent Similarity: 48.921 Percent Identity: 20.863

alignment_block:

US-09-701-711-2 x AAX91713

Align seg 1/1 to: AAX91713 from: 1 to: 2673

```

5 TyrPheLysGlyPheGlnValSerAlaMetThrMetAlaValMetMetVa 21
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4 TACAAAAGATTATTCGCGTAGCAGCTCTCTTCGCGCAGCATAGGGAT 53
21 MetSerThrHisAlaGlnAlaAspPheMetAlaAsnAspIle.... 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
54 CCGTGA....GGACAGTCTCGGATCGACCTGACCCCTCAGGATACATAT 97
37 .....
98 ATAGCCCTGAATATCCATGATGCCAAGCCCTATTCATAAGACCATGCTCT 147
41 LeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheArgLe 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 ATTGAG.....ATCGAGGATAGGCTTTTCGATGACTTTGTCTT 188
57 u.....GlyGlnValValSerGluAsnGlnLeuAlaAspGly. 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 GCGCAATCTTTTCAGGCTTGGCTGTAGGTGATGAAGTCTGATTCCTGGAG 238
70 .....VallysAlaLeuTyrAlaThrGlyAsnPheSer 80
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
239 ATGCCATCTGCTGCTGCCGTGAATGAATATTCGCGTACGGCTACTCTCA 288
81 AspValGlnValTyrHisGlnGlu.....GlyArgIleIleTyrGln... 94
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
289 AATGTGCAATCATCGCGGATAAATATTCGCGCAATAAGTCTATCTGAA 338
95 .....ValThrGluArgProIleAlaGluIleAsnPheGluGlyA 109
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
339 AATCATTTCTCACTGAACCTCTCGCATCAGTAAGGTACTTTTAGCGGG 388
109 snArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAlaGly 125
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 TAAGAACTCTGAGAGAGATCTTGAA.....ATGAAATCGGT 429
126 LeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluTh 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
430 CTTCCGAGGGGATTTCAGATCACCAGAAATAATGAAGACAAAGTTCAGCA 479
142 rGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleT 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
480 AATCGTACAGAAGTATTTTAGTGAGAAAGTTATCGCGATGCCAGCATAC 529
159 hrValLysGln.....ThrMetLeuAspGlyAsnArgValLysLeuAsp 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 GGATTAACCAAGCAACCGGATCTTTCCAAAGATGGCTTTGTCATGTGCTT 579
174 MetThrPheAlaGluGlyLysProAlaArgValValAspIleIleI 190
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
580 ATCTCGATTGAGAGAAAGCAAAACCAAGGTGAATGAATTTATTTTC 629
190 eGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuAlaIleL 207
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
630 CGGCAACCAAGCCCTTAGCAATCATAAAGCTA.....AGAATGGCGATGA 673
207 ysAspAsnLysIleAsn..... 212
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
674 AGAACCCCAATGCCCAATTCAGTCTTAGAAGCATATTCGCTCATCTTTC 723
213 ...ProLeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSe 228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
724 TTGAAACTTTTATGACTACATCAAGTTTCTGGAGAGAGACTACCGTGAAGA 773
228 rLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheG 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

774 TTTGGTCCGATTGATAGAGAAGTATCAGGAATATGATATCGTGATGCTG 823
245 luileLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePhe 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
824 AATACTGACCGACAGTGTCTGGAAGGCTCTGACGCGCAAGAGTGGAT 873
262 ValGluIleSerLeuHisGluGlyGluGlnTyrArgPheGlyGlnThrGl 278
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874 ATTATCTCAACATCGAAGAGGGCAGAGATATATTAAGGATGTCAA 923
278 nPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuL 295
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
924 CTTTGTGGCAATTCAATATATCCATCGGAGTATTTGGAACGAGTCTGCG 973
295 ysPheLysAlaGluGlyPheSerGlnAlaMetLeuGluGlnThrThr 311
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
974 GAATAAATCCGAGATGTACAATCAGAGAGGATTGGCTAAGCGTCTC 1023
312 AsnAsn.....IleSerThrLysPheGlyAspGlyTyrTy 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1024 AATGAAGATGAAGTCTCTGCGGACCTGTACTATAACAATGGCTATAT 1073
324 rTyrAlaGlnIleArgProVal...ThrArgIleAsnAspGluSerArgT 340
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1074 TTTTGGTGGTCTCGATCCGCTGGAACAAATGTAGTGGGGATCTGTTT 1123
340 hrValAspValGluTyrTyrIleAspProValHisProValTyrValArg 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1124 CCCTTGATATTCGT.....ATAGCGGAGGGGAAGCAGGCCAATATCAAT 1167
357 ArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgAr 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1168 AAGGTGATCATCAAGGAATACTGTCGTGTACGAAGACGTAGTACGCG 1217
373 gGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnL 390
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1218 AGAGCTTTACACAAAGCCGCGCAGCTCTTTAGTCGCGAGGATATCATTA 1267
390 euSerArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrVal 406
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1268 ACTCATTCGCTCATCAATCAGTTGGGCATTTTCATGCCGAAATCT 1317
407 AspThrArgProValProAsn...SerProAspGlnValAspValAsnPh 422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1318 ATTCGCCGTCGCGATTCCCAATCCGAAACAGGACAGTGGATAGAGTA 1367
422 eValValGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrS 439
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1368 TGATTTGGTCCGCGTAGCAGTACCAATTTGGAGCTTTCTGCGGTGGA 1417
439 erGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsnAsnPh 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1418 GTCACTCC...GGACTTCTGTTCCGAGGAGCCATTAAGTTTCAGCAACTTC 1464
455 ..... 455
1465 TCTGTGCGCAACTTGCTCCATCCCTCGATGTATAAGAAAGGGATCATTC 1514
456 MetGlyThrGlyLysHisValAsnAlaSerPhe...SerArgSerGluT 471
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1515 GCAAGGGATGGCAACACATATCATCTAGTGTCTCAGACCAATGGAAGT 1564
471 hrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheThrValAsn 487
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1565 ACTATCAGCAGTATATGTGCACATTTATGTCATCCATGGTTTGGGGCAAG 1614
488 GlyValSerGln...SerLeuSerGlyTyrTyrArgLysThr..... 500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1615 CGGCGGATATGTTTCAGCTTCAGTGTCTTATTTCCAAAGACTACGCGGAT 1664
501 .....LysTyrAspAsnLysAsnIleSerAsnTyrVal..... 511
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1665 TGACTCCAGTTCATCAATAGCAATGCGCGCAACTACTATAATGCCTACT 1714

```

```

512 .....LeuAspSer 514
1715 ATAATAGCTACTACAACAATAATAATAGTTATTACAACGGTATATCGAAC 1764
515 TyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIle 531
1765 TATACCGCGGACCTTC...TATACCTACGCGCAGCGATCCGGATCGTTCGCT 1811
531 eserPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheM 548
1812 T.....CAGATGT 1819
548 etGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGln 564
1820 TAGGTACTTCG.....ATCGGTTACGGTAAGCGTTTGACT 1854
565 ValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAs 581
1855 TGCCCGGACAAATGGTTCAGATTTATACCTCTCTGAATACACCTACTA 1904
581 nAlaIleLeuGlyTrpAsnTyr..... 588
1905 TAGACTCGGAATTTGGAGCTACAATACCTTCCAAATTTCCATCATGGCT 1954
589 .....SerSerLeuAsp 592
1955 CGGCTAATGATCTCAACTTGGAGCTGCGTCTCTCTCGTACTTCCATCGAT 2004
593 ArgProValPheProThrGlnGlyMetSerHisSerValAspLeu..... 607
2005 AATCCTATTATACCAAGACGGATCGGATTCATGTTTCTGTGCTGC 2054
608 .....ThrV 609
2055 TACTCTTCCTTATTCTTTGGGACAATCATGACTATGCCAGCCAGAAC 2104
609 aGlyPheGlyAspLys.....ThrHisGlnLysValValTyr 621
2105 TCAGCGTAAGCGATCGTTACAGATTTATCGAGTATCACAAAGTGAAGTTT 2154
622 GlnGlyAsnIleTyrArgProPheIleLys..... 631
2155 AGAGGAGGAGTTTACTTCCATTCGTCATCCGCTACGCATAAATATAC 2204
632 .....LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyA 645
2205 ACCGGTCTCATGAGTCGAGTGAAGGAGCAGTTCTTGGTTCGTATAATT 2254
645 sAsn.....LeuProPheTyrGluAsnPheTyrAlaGlyGly..... 657
2255 CCAATAAGAAATCTCTTTC...GGTACTTCTATATGGGAGGTGATGGT 2301
658 .....TyrGlySerValArgGlyTyrAspGlnSerSerLeuGly 670
2302 ATGTCAGCATATTATGGT.....GGCTACATGAATGAGACTATAGG 2342
670 yProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnGlnThrL 687
2343 T.....TTGCGTGGTTATAAGAAC... 2361
687 euGlyGlyValValGlyLysAsn.....AlaLeuAlaThr 698
2362 ..GGATCTATTCGGGTATAAATACTAGACTATGCATATGCATATATGCGG 2409
699 PheGlySerGluLeuIleLeuProLeuProLysGlyAsp..... 712
2410 CTTACGATGAAGTACGTTTCCGATTCGTTTGAACAATCATTCATGC 2459
713 .TrpIleAspGlnValArgProValIlePheLeuGlyGlyGlnValP 729
2460 GTGGCTC.....TTAGCTTTTCCGAAGCAGGCAATCGCT 2494

```

```

729 heAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLys 745
2495 GG.....CGCAGTATCGAC..... 2508
746 AspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLe 762
2509 .....AATATAATCCCTTTAAC..... 2526
762 uLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrT 779
2527 .....CTGAAGCGATCGCCGGGTAGGATTCGCTG 2557
779 rpTyrThrPro...IleGlyProLeuSerIleSerTyrAla..... 791
2558 TAAGCTTACCGATGTCGGAATGTCGATATCGATTCGGGATATGGCTTT 2607
792 ...LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPh 807
2608 GACCGTCCGGACAATCTCTACAGCGAGGAGGAAGCAAT...GTCCACTT 2654
807 eGlnIleGlySerValPhe 813
2655 TGTCTCGACAGGATTC 2673
seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AAZ33105
seq_documentation_block:
ID AAZ33105 standard; DNA; 2514 BP.
XX
AC AAZ33105;
XX
DT 26-JAN-2000 (first entry)
XX
DE Treponema pallidum D15/Oma87 homologue DNA.
XX
KW Treponema pallidum; syphilis; vaccine; treponemal disease; Msp;
KW major sheath protein; bejal; gingivitis; periodontal disease; pinta;
KW yaws; PCR primer; ss.
XX
OS Treponema pallidum.
XX
PN W09953099-Al.
XX
PD 21-OCT-1999.
XX
PF 09-APR-1999; 99WO-US07886.
XX
PR 10-APR-1998; 98US-0058968.
XX
PA (UNIW ) UNIV WASHINGTON.
XX
PI Van Voorhis WC, Lukehart SA, Centurion-Lara GA, Cameron CES;
XX
WPI: 1999-620445/53.
XX
P-PSDB; AAY52774.
XX
PT Novel proteins useful in vaccines against syphilis and other treponemal
XX diseases -
XX
PS Claim 2; Page 95-99; 200pp; English.
XX
CC The present invention describes novel Treponema pallidum genes. These
XX genes encode a glycerophosphodiester phosphodiesterase (Gpd), a
XX D15/Oma87 homologue, and proteins with homology to major outer sheath
XX (Msp) proteins of T. denticola. Also described are: (1) an isolated
XX protein capable of inducing a protective immunologic response to
XX T. p. pallidum, T. p. pertenuis, or T. p. endemicum, when administered
XX in an effective amount to an animal host; (2) a method of inducing
XX a T. p. pallidum vaccine candidate; (3) a method of inducing a
XX protective immune response against T. pallidum; (4) a method for
XX analysing a sample of DNA to determine whether it originated from
XX T. p. subspecies pallidum, T. p. subspecies pertenuis or T. p. subspecies
XX endemicum; and (5) a method of determining whether a first and a second

```



```

504 nLysAsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSert 521
1521 CAATAATCTCTTTGTGTACCGCGCGGTTCATACGCAACGGGTCCGC 1570
521 yGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAla 537
1571 ACCGTTACAGCAGCAGGAGCAGTGGGCTAGTCCCTCGGCTG...GCA 1617
538 AspAsnThrLysLeuHisGlyIleArgPhe.....MetGlyIleSe 551
1618 GAATCGTTCCGCTCAAGTATTCCGCGTTTCAGTCCCGCCATCGCGCGCA 1667
551 rAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspAsnA 568
1668 CACCGGGTACCATCGTGTATCCGCGCTATCGCGTCATTAGGCTG.....A 1711
568 sNGlyIleProAspPheLys..... 574
1712 ACGGGGGGTGGACTTTCGGGTGTAAAGAAATTTTACGATAAGGATAAC 1761
575 .....HisAspTyrThrThrTy 580
1762 AATCAGCCCTTCGACCTGACCGTAAAGACGACGTGAACCTGGACAGTAT 1811
580 rAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgProValPheP 597
1812 CAATTCGTTT.....TGGACGACGTTTCGTTTGAC..... 1842
597 roThrGlnGlyMetSerHisSerValAspLeuThrValGlyPheGlyAsp 613
1843 .....GGCGTGTACTTGGCTACGACCGCTCCAGCGCTGG..... 1878
614 LysThrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIl 630
1879 TTTTATAGCAGCGCTGTACGTTCACGCGCTCGTT.....CCCTTCT 1922
630 eLysLysSer...ValLeuArgGlyTyrAlaLys..... 640
1923 CGAAAAGAGCATTCGTTTCGCTCGACACCAAGCGCGCTTCTACGTTA 1972
641 .....LeuGlyTyrGlyAsn.....AsnLeuProPheTyrGlu 651
1973 CCTGCTCAATTATCGGCTCTCGCGTGGGAACTTAAGTTTGTCTTG 2022
652 AsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSerSe 668
2023 GCTTCTACACCGGT.....GTGTCGTTCAAACGCTAT..... 2055
668 rLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnGlnT 685
2056 .....TATGGACGGAGGAAAGCGAAA 2077
685 hrThrLeuGlyGluValValGlyAsnAlaLeuAlaThrPheGly... 700
2078 ACGGAAAGGCAACGGGTGCGGTGCGCGCTGTAATAGACGCGTG 2127
701 .....SerGlute 703
2128 CTGGTAGGCGCGGTGGAGCAAGACGCAAAAGAAAACACCGGAGACCT 2177
703 uileLeuProLeuProPheLysGlyAspTyrIleAspGlnValArgProV 720
2178 GCTGCTC.....CACCATTGGATTGAGTTCGCTCCGCTCGCGCC 2212
720 al.....IlePheIleGluGlyGlnVal 728
2213 TGGCGCAGCGCATTCGTCTTTCGCTTTCCTTCATCGCGCAATGCTG 2262
729 PheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnPheLy 745
2263 TACACATCGAA.....AGTCACGTCCCC 2285
745 sAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProL 762

```

```

...
2286 AAACGGGTCTATCGTCCCGCCAGCAGCTCCAGCAGCAGTAGTAGTAGCA 2335
762 euleuThrGlnAspLysGlnLeuArgTyr.....SerAlaGlyValGly 776
2336 GTAGAACCACCACTCTGAAGGACTGTACAAATAGACTACGCTCCGGG 2385
777 AlaThrTyrTyrThrProIleGlyProLeuSerIleSerTyrAlaLys... 792
2386 CTGGGCTTTACATTCGCGCAATTCCTGTTAAATTCGCTTCGCAACAC 2435
793 .....ProLeuAsnLysLys 797
2436 CTTACGTCACCGCGCGCATCCCAAAACAAAGAA 2472

```

seq_name: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:AA233106

seq_documentation_block:
ID AA233106 standard; DNA; 2439 BP.

```

XX AC AA233106;
XX XX
XX DF 26-JAN-2000 (first entry)
XX DE Treponema pallidum D15/Oma87 homologue partial DNA sequence.
XX KW Treponema pallidum; syphilis; vaccine; treponemal disease; Msp;
XX KW major sheath protein; bejal; gingivitis; periodontal disease; pinta;
XX KW yaws; PCR primer; ss.
XX OS Treponema pallidum.
XX PN W09953099-A1.
XX XX
XX PD 21-OCT-1999.
XX PF 09-APR-1999; 99WO-US07886.
XX PR 10-APR-1998; 98US-0058968.
XX PA (UNIW ) UNIV WASHINGTON.
XX PI Van Voorhis WC, Lukehart SA, Centurion-Lara GA, Cameron CES;
XX DR WPI; 1999-620445/53.
XX DR P-PSDB; AAY52775.
XX PT Novel proteins useful in vaccines against syphilis and other treponemal
XX PT diseases
XX PS Claim 2; Page 102-105; 200pp; English.
XX XX

```

The present invention describes novel Treponema pallidum genes. These genes encode a glycerophosphodiester phosphodiesterase (Gpd), a D15/Oma87 homologue, and proteins with homology to major outer sheath (Msp) proteins of T. denticola. Also described are: (1) an isolated protein capable of inducing a protective immunologic response to T. p. pallidum, T. p. pertenue, or T. p. endemicum, when administered in an effective amount to an animal host; (2) a method of inducing a protective immune response against T. pallidum; (3) a method of inducing a protective immune response against T. pallidum; (4) a method for analysing a sample of DNA to determine whether it originated from T. p. subspecies pallidum, T. p. subspecies pertenue or T. p. subspecies endemicum; and (5) a method of determining whether a first and a second clinical isolate of T. p. pallidum are the same or different. The proteins are used, either alone or in combination, in vaccines against Treponemal diseases, e.g. syphilis, bejal, pinta, yaws, gingivitis, and periodontal disease. They may also provide protection against other Treponemal diseases. The methods may be used to identify vaccine candidates, and to determine the origin of a treponemal nucleic acid. AA233104 to AA233156, and AAY52773 to AAY52831, represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 2439 BP; 596 A; 544 C; 731 G; 568 T; 0 other;

alignment scores:

Quality:	348.50	Length:	847
Ratio:	0.783	Gaps:	28
Similarity:	52.538	Percent Identity:	21.133

alignment block:

US-09-701-711-2 x AAZ33106

Align seq 1/1 to: AAZ33106 from: 1 to: 2439

[illegible]

275 GlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAlaGluLeuG1 291
||| : : : : : ||||| : : : : : ||||| : : : : :
760 GCGGGGTTTACCATTGTGGTAAACAGATTTTTATGACACCGAGGAGCTGCA 809
291 AlaLeuLeuGlyPheLysAlaGluGluGlyPheSerGlnAlaMetLeuG 308
||| : : : : : ||| : : : : : : : : : : : : : : :
810 GGC AAAAATTAGGCTCAAGCGCGGGCCATCATGAATATGCTGGCCCTTG 859
308 LuGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyr 324
||| :
860 AGCAGGCTTTCAGCGCTGCGGATCGGTATTTGAAACAGGATACACG 909
325 TyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgThrVa 341
||| :
910 TC AAAATTACCTGAACAAGAAGAACAACGGGACACGCGCAGAAAACGCT 959
341 LaspValGluTyrTyrIleAspProValHisProValTyrValArgArgI 358
||| :
960 TTCGTTTAAGATACGGTGTGGAGCGGACGCGCAGCGCATCGAGCACA 1009
358 LeAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgGlu 374
||| :
1010 TTATCAATTAGGAAACGAAGAATAACAACGACGAGTTATCTCCGCTGAA 1059
375 MetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSe 391
||| : : : : : ||| : : : : : : : : : : : : : : :
1060 ATGCTGCTGAACCGGGGATGTCTTCTCAAGTCAAGTTTACGGATAG 1109
391 rArgAlaArgLeuMetArgThrGlyPhePheLysHisValThrValAspT 408
||| :
1110 CTGCGCAATCTGTTCAACCTGCCGTATTCTCGCTCGTGGCGGATG 1159
408 hrArgProValProAsnSerProAspGlnValAspValAsnPheValVa 124
||| : : : : : : : : : : ||| : : : : : : : : : :
1160 TCGGCCCC...GGCTCTCAGCAGGACCTGGTGGACATTATCTCCAATGTG 1206
425 GluGluGlnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSe 441
||| :
1207 GAGGAGCAGTCCAGCGGAAACGTCAGTTTGGGTGACGTTTCTGGGT 1256
441 rGlyGlyVal...ThrPheGlnPheAspVal...SerGlnA 453
||| : : : : : ||||| : : : : : : : : : : : : : : :
1257 GGGGAGCAGCAGGACGTTCCGCTTTCCGCTTTTGCATGGGAAGAAA 1306
453 snAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArgSer 469
||| : : : : : ||| : : : : : : : : : : : : : : :
1307 AGAATTTTTCGGAAAGGGAATGAATTTTCAGTAATGCAACCTGGGG 1356
470 GluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheThrVa 486
||| :
1357 TCTGAGCGCAGACCTGAACTCGGTATGTGGAGCGCTGTTTCTGGG 1406
486 LasnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysTyrA 503
||| :
1407 CTCCTCCGCTGACGTGGGCTTTGAC...TTTGAACCTTA 1441
503 spAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeu 519
||| : : : : : ||| : : : : : : : : : : : : : : :
1442 GCACAAAAATCTCTTTGTGTACCGCGGGGTCATACGCAACGGGCTG 1491
520 SerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAs 536
||| : : : : : ||| : : : : : : : : : : : : : : :
1492 CCGCAGCGTACACGACGAGGACAGTGGCTAGTTCCCTCGGGCTG... 1519
536 nAlaAspAsnThrLysLeuHisGlyGlyArgPhe...MetGlyI 550
||| :
1540 .GCAGATCGTTTCGCTCAAGTATTTCGCTTTGAGTCGCGCATCGGCG 1588
550 LeSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAsp 566
||| : : : : : ||| : : : : : : : : : : : : : : :
1589 CGCACCGGTCACCATGGTATCCCGGCTATCGGTCATTAGGGTG... 1635
567 AsnAsnGlyIleProAspPheLys... 574

```
1636 ..... ||||| ||||| .....
575 ..... HisAspTyrThrT 579
1683 TAACATCAGCCCTTCAGCTGACCTGACCTGAAAGAGAGAGCTGAAGTGGACCA 1732
579 hrTyAsnAlaLeuLeuGlyTyrAsnTyrSerSerLeuAspArgProVal 595
1733 GTATCAATTCGTTT ..... TGACGAGCGTTCGTTTGAC ..... 1767
596 PheProThrGlnGlyMetSerHisSerValaspLeuThrValGlyPheL 612
1768 ..... GGGCGTGACCTTCGTCAGCACCCGCTCCAGCGGCTGG... 1803
612 yAspLysThrHisGlnLysValValGlnGlyAsnIleTyrArgProp 629
1804 ..... TTTTAGCAGCGCTGTACGTTCAACGGGCTCGTT ..... CCCT 1843
629 heLleLysLysSer... ValLeuArgGlyTyrAlaLys ..... 640
1844 TTCTGAAAGAGCATTCGTTCTCGTCGACACCAAGCGCGAGTCTAC 1893
641 ..... LeuGlyTyrGlyAsn ..... AsnLeuProPheTyr 650
1894 GTTACCTGCTCAATATCCGCTCTCTCGCGGTGGAACTTAAAGTTTGT 1943
650 rGluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnS 667
1944 CTTGGCTTCTACACCGGT ..... GTGCTCGTTCAACCGTAT ..... 1980
667 erSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGln 683
1981 ..... TATGGACGGGAGAAAGC 1998
684 GlnThrThrLeuGlyGluValValGlyGlyAsnAlaLeuAlaThrPheL 700
1999 GAAACGGAAGGCAACGGGTGGCGTCCGGCGCGTGGTAATAGACGG 2048
700 y ..... SerG 702
2049 CGTGCTGTAGGCGCGGTGGAGGAGAGAGCAAGAAAGAAACACCGGAG 2098
702 luLeuIleLeuProLeuProPheLysGlyAspTyrPheAspGlnValArg 718
2099 ACCTGCTGCTC ..... CACCACTGGATTGAGTTCGCTGG 2133
719 ProVal ..... IlePheLleGluGlyGlyL 727
2134 CGGCTGGCGCAGCGCATGTGCTCTTTCGACTTTTCTTTGATGCGCAAT 2183
727 nValPheAspThrThrGlyMetAspLysGlnThrIleAspLeuThrGlnP 744
2184 GGTGTACAAATCGAA ..... AGTCAGT 2206
744 heLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArg 760
2207 CCCCAAGCGGTATCGTCCCGCAGCAGCTCCAGCAGCAGCAGTAGTAGT 2256
761 ProLeuLeuThrGlnAspLysGlnLeuArgTyr ..... SerAlaGlyVa 775
2257 AGCAGTAGAACACCAACGAGCTCTGAAGGAGCTGTACAAAATGAGCTACGGTCC 2306
775 lGlyAlaThrTyrThrProLleGlyProLeuSerIleSerTyrAlaL 792
2307 GGGCTGGCTTTACATTGCCGCAATTCGTTAAATTTGGCGTTCGCAA 2356
792 ys ..... ProLeuAsnLysLys 797
2357 ACACCTTCACGTCACCGGGCGGCATCCCAAAACAAAGAAA 2397
```

seq_name: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:AAV25052

seq_documentation_block:
ID AAV25052 standard; DNA; 1845 BP.

XX AC AAV25052;

XX DT 24-JUN-1998 (first entry)

XX DE H. pylori cell envelope OMP ORF 05ep10815_4719175_c1_83.

XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
KW identification; binding compound; bacteria; life cycle; activator;
KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis; ds.
XX OS Helicobacter pylori.XX FH Key Location/Qualifiers
FT CDS 1..1845
FT /*tag= a

XX PN WO9737044-A1.

XX PD 09-OCT-1997.

XX PF 27-MAR-1997; 97WO-US05223.

XX PR 06-DEC-1996; 96US-0761318.

XX PR 29-MAR-1996; 96US-0625811.

XX PR 02-APR-1996; 96US-0758731.

XX PR 25-OCT-1996; 96US-0736905.

XX PR 28-OCT-1996; 96US-0738859.

XX PA (ASTR) ASTRA AB.

XX PI Alm RA, Smith D;

XX DR WPI; 1997-503122/46.

XX DR P-PSDB; AAW55643.

XX PS Claims 5,6,21; Pages 417-418; 1145pp; English.

This DNA is stated to encode a H. pylori cell envelope outer membrane protein having a terminal Phe residue. The protein may be used in a vaccine to prevent or treat H. pylori infection or to identify H. pylori polypeptide binding compounds, useful as potential H. pylori life cycle activators or inhibitors. The DNA and probes derived from it may be used for the identification of H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic acid sequences complementary to the DNA act as antisense sequences and can be used to prevent the translation of H. pylori mRNA. Antibodies against the protein can be used in immunoassays to evaluate the abundance and distribution of H. pylori-specific antigens. The genomic sequence of H. pylori (ATCC 55679) was determined from overlapping contigs generated by mechanically shearing the bacterial DNA. The sequences were analysed for ORF of at least 180 nucleotides, and the predicted coding regions defined by computer evaluation. To identify likely H. pylori antigens for vaccine development, the amino acid sequences predicted from various ORF were analysed for significant homology to other known or exported membrane proteins. Having identified and determined the sequences of interest, particular regions can be isolated from H. pylori by PCR amplification for recombinant polypeptide production, e.g. in E. coli hosts.

Sequence 1845 BP; 560 A; 329 C; 473 G; 483 T; 0 other;

alignment_scores:

Quality: 344.50

Ratio: 1.101

Length: 557

Gaps: 17

Percent Similarity: 56.194 Percent Identity: 23.339

alignment_block:

US-09-701-711-2 x AAV25052

Align seg 1/1 to: AAV25052 from: 1 to: 1845

```
30 AspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIle 46
   :::::::::::::::::::::
85 GAAATGAAAGTCAAGTCCATTTCTTATGTCGGGCTTTCTTACATGCTGA 134
   :::::::::::::::::::::
46 eGluSerLeuGlnSerValLeuProPheArgLeuGlyGlnValValSerG 63
   :::::::::::::::::::::
135 CATGTCGCTATGAATTCGAAGATTCGCGGGCGATATGCTGGAT 184
   :::::::::::::::::::::
63 LuAsnGlnLeuAlaAspGlyValValValLeuTyrAlaThrGlyAsnPhe 79
   :::::::::::::::::::::
185 CTAAATAAATAGACACCGCTGTTTGTGCTTCAACCAAGGCTATTT 234
   :::::::::::::::::::::
80 SerAspValGlnValTyrHisGlnGluGlyArgIleIleTyrGlnValTh 96
   :::::::::::::::::::::
235 AAAGACGTTTATGCCACTTTTGAACACGGCATTTTAGAGTTTCATTTGA 284
   :::::::::::::::::::::
96 rGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleP 113
   :::::::::::::::::::::
285 TGAAGAAGCCAGGATTCGCGGGTGGAATCAAGGT.....TAGG 325
   :::::::::::::::::::::
113 rOlyGluGlyLeuGlnGluGlyLeuLysAsn...AlaGlyLeuAlaVal 128
   :::::::::::::::::::::
326 GGACTGAAAGAAAGAAAGACGGCTTAAATCCCAATGGGGATCAAAAG 375
   :::::::::::::::::::::
129 GlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuTh 145
   :::::::::::::::::::::
376 GCGCACACCTTTGATGAGCAAAATTAGACATGCTTAAACGGCTTAAA 425
   :::::::::::::::::::::
145 rAsnGlnTyrIleSerGlnGlyTyrAsnThrGluIleThrValLysG 162
   :::::::::::::::::::::
426 AACGGCTTAGAGGGCGAGGCTATTATGGAGCGTGGTGGAGGTCGCCA 475
   :::::::::::::::::::::
162 lnThrMetLeuAsp..... 166
   :::::::::::::::::::::
476 CAGAAAGCTCAGTGAGGAGCGCTTATTGATCGTGTGATGATAGG 525
   :::::::::::::::::::::
167 GlyAsnArgValLysLeuAspMetThrPheAlaGluGly.....LysPr 181
   :::::::::::::::::::::
526 GGGGACAGTATTATATCAACAACATCCATTTATGAGGAGCGGATAAAT 575
   :::::::::::::::::::::
181 oAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSerAspA 198
   :::::::::::::::::::::
576 AAACCCCGTGTGATGATGATCTTTGAGCGCGAACAG.....CAGC 616
   :::::::::::::::::::::
198 laAspLeuIleAsp...ValLeuAlaIleLysAspAsnLysIleAsnPro 213
   :::::::::::::::::::::
617 GCGATTTTCATGGCTGGATGTCGGGCTTGAAATCAGCGGAATG..... 660
   :::::::::::::::::::::
214 LeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGl 230
   :::::::::::::::::::::
661 .....CGCTTAGATCAATTAGAATACGATCTTTGCGTATCCAGATGT 704
   :::::::::::::::::::::
230 uAsnLeuArgAlaLysTyrLeuAsnAla.....GlyPheValA 243
   :::::::::::::::::::::
705 GTATATGCGTAGGGGTACTTACGCTCATATTTCTCGCCTTTTGTGA 754
   :::::::::::::::::::::
243 rg.....PheGluIleLysAlaLysLeuAsnIleAsnGluAspLys 257
   :::::::::::::::::::::
755 AAACGGATTTTCCACCCATGACGCTAAGTCCAT..... 789
   :::::::::::::::::::::
258 AsnArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPh 274
   :::::::::::::::::::::
790 .....TATAAGGTCAAAAGAGGGGATCCCAATACAGGAT 821
   :::::::::::::::::::::
274 eGlyGlnThrGlnPhe...LeuGlyAsnLeuThrTyrThrGlnAlaGluL 290
   :::::::::::::::::::::
822 TTCAGATATTTTAATAGAGATTGACAACCCGGTAGTCCCTTAAAAACCT 871
   :::::::::::::::::::::
```

```
290 euGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMet 306
   :::::::::::::::::::::
872 TAGAAAAAGCGCTTAAAGTTAAAGGAAGAGATGCTTTAATATTAGCAT 921
   :::::::::::::::::::::
307 LeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTy 323
   :::::::::::::::::::::
922 TTAAGAGCGGATCGCAAAATTTTAAAAACCGAAATCGCGATAGGGCTA 971
   :::::::::::::::::::::
323 rTyrTyrAlaGlnIleArgProValThrArgIleAsnAspGluSerArgT 340
   :::::::::::::::::::::
972 TCGCTTTCGGTGGTGAAGCCAGACTTGGATAAAGACGAAAAAACGGGC 1021
   :::::::::::::::::::::
340 hrValAspValGluTyrTyrIleAspProValHisProValTyrValArg 356
   :::::::::::::::::::::
1022 TTCTGAAAGTCATTATTATCGTATTGAAGTGGCGATATGTTGCATATCA 1071
   :::::::::::::::::::::
357 ArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgAr 373
   :::::::::::::::::::::
1072 GATGTCATCATTTTCAGGGAACCGACGACGAGCATAGGATCATTAGGAG 1121
   :::::::::::::::::::::
373 gGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnL 390
   :::::::::::::::::::::
1122 GGAATG.....TTACTAGGCGCTAAAGATAAATACAACT 1156
   :::::::::::::::::::::
390 euSerArgAlaArg.....LeuMetArgThrGlyPhePhe 401
   :::::::::::::::::::::
1157 TGACCAAACTGAGAAATTCGAAAAATTTCTTTGAGCGCTTTAGGCTTTTC 1206
   :::::::::::::::::::::
402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418
   :::::::::::::::::::::
1207 TCTAAAGTCAAGATTGAAGAAAAAAGGGTCAATAGCTCA.....TTGAT 1250
   :::::::::::::::::::::
418 laSpValAsnPheValValGluGlnProSerGlySerSerThrIleA 435
   :::::::::::::::::::::
1251 GGATTTCTTAGTGAGCGTAGAAGAGGGCGCCGCGCATGTCGAATTCG 1300
   :::::::::::::::::::::
435 laAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspValSer 451
   :::::::::::::::::::::
1301 GGTGGCTATGCTCTTATGAGGGCTCATGCTTAATGGAGCGTGAGC 1350
   :::::::::::::::::::::
452 GlnAsnAsnPheMetGlyThrGlyLysHisValAsn.....AlaSerPh 466
   :::::::::::::::::::::
1351 GAAAGGAATCTTTTGGCAGGCAAAAGCATGAGCTTGATGCTAACAT 1400
   :::::::::::::::::::::
466 eSerArgSerGluThrArgGlu..... 473
   :::::::::::::::::::::
1401 TGCCACAGGGGGGTAGATCTTATCGGGCATGCCAAAAGGGCGGGGC 1450
   :::::::::::::::::::::
474 .....ValTyrSerLeuGlyMetThrAsnProTyrPheThrValAsn 487
   :::::::::::::::::::::
1451 GTATGTTTTCGGGGAATTTGAGCTTGACTAATCCAAGGATTTTGGACAGC 1500
   :::::::::::::::::::::
488 GlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysTyrAspAs 504
   :::::::::::::::::::::
1501 TGGTATAGCTCTACGATCAATCTTTAT.....GCGGATTA 1535
   :::::::::::::::::::::
504 nLysAsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerT 521
   :::::::::::::::::::::
1536 CAGATAAGCTACCAATACATACCAAGCGGGGGCTTTGGGTGAATG 1585
   :::::::::::::::::::::
521 yrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAla 537
   :::::::::::::::::::::
1586 TCGGGCGCATGCTGGGTAAATAGAACCCATGTGAGCTTAGGTATAACTTG 1635
   :::::::::::::::::::::
538 AspAsnThrLysLeuHisGly 544
   :::::::::::::::::::::
1636 AATGTTACCAAACTCCTTGGT 1656
   :::::::::::::::::::::
```



```

/cgn2_6/ptodata/2/ina/backfiles1.seq:5169629-1 + 130.00 215.75 0.0002 3531 1
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-418-893D-25 + 129.50 207.07 0.0008 627C
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-613-009A-4 + 128.50 219.54 0.0002 2106
/cgn2_6/ptodata/2/ina/6A.COMB.seq:US-08-613-009A-3 + 128.50 218.67 0.0002 2247
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-841-178-26 + 128.50 212.64 0.0004 3534

```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-433-522A-1

```

seq_documentation_block:
; Sequence 1, Application US/08433522A
; Patent No. 6013514
;
; GENERAL INFORMATION:
;
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Two Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
;

```

; CURRESPONDENCE ADDRESS:
; ADDRESSEE: SIM & McBURNEY

ADDRESSEE: SIM & MCBURNEY
STREET: 6TH FLOOR 330 UNIVERSITY AVENUE

STREET: 6TH E
CITY: TORONTO

CITY: Toronto
STATE: Ontario

STATE: Ontario
COUNTRY: Canada

; COUNTRY: Canada
; ZTD: DEC 197

```

; ZIP: M5G 1R7
; COMPTON PRESS

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1 0. Version #1 25

```

```

? SOFTWARE: PatentIn Release #1.0, version #1.25
?
? CURRENT APPLICATION DATA:
?
? APPLICATION NUMBER: US/08/433,522A
?
? FILING DATE: 12-SEP-1995
?
? CLASSIFICATION: 435
?
? ATTORNEY/AGENT INFORMATION:
?
? NAME: STEWART, Michael I.
?
? REGISTRATION NUMBER: 24,973
?
? REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
?
? TELECOMMUNICATION INFORMATION:
?
? TELEPHONE: (416) 595-1155
?
? TELEFAX: (416) 595-1163
?
? INFORMATION FOR SEQ ID NO: 1:
?
? SEQUENCE CHARACTERISTICS:
?
? LENGTH: 2949 base pairs
?
? TYPE: nucleic acid
?
? STRANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: DNA (genomic)
?
? FEATURE:
?
? NAME/KEY: CDS
?
? LOCATION: 75..2465
?
? US-08-433-522A-1
?
alignment_scores:
?
? Quality: 1052.50 Length: 822
?
? Ratio: 1.956 Gaps: 20
?
? Percent Similarity: 65.450 Percent Identity: 31.630
?
alignment_block:
?
? US-09-701-711-2 x US-08-433-522A-1 ..
?
Align seg 1/1 to: US-08-433-522A-1 from: 1 to: 2949
?
23 SerThrHisAlaGlnAlaAAspPheMetAlaAsnAspIleThrIleTh 39
?
? ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
?
114 ACAACGACGTGTGTTGCCGACACCTTTTGTGCCAAAGATATTCGTGGA 163
?
? ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
?
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
?
? ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
?
164 TGGTGTGTTCAAGGTGACTTACAGAACAAATCCGAGCAAGTTTACCTGTC 213
?
? ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

56 rGleuGlycInValValSerGluasnGlnLeualaAspGlyVallysAla 72
214 GTGCGCGGTGACGCGTGTACTGACATGATGTGGCGTAATATGTTCGCGTCT 263
73 LeuTyRAlaThrGlyAsnPheSerAspValGlnValTyRHisGlnGluGl 89
264 TTATTTCGTAAGTGGCGAATTGATGATGTGAACGG...CATCAAGAAGG 310
89 yArgile...IleTyRGlnValThrGluArgProLeuIleAlaGluIleA 105
311 CGATGCTGTTGTTGTAGCGTGTGGCTAAATCGATCATTTTCAGATGTTA 360
105 snPheGluGlyAsnArgLeuLeuProLysGluGlyLeuGlnGluGlyLeu 121
361 AAATCAAAAGGTAACTCTGTTATTCCTACCTGAAGCAGCTAAACAAAACCTT 410
122 LysAsnAlaGlyLeuAlaValGlyClnProLeuLysGlnAlaThrValGl 138
411 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAATTTAA 460
138 nMetIleGluThrGluLeuThrAsnGlnTyRILESerGlnGlyTyRtyrA 155
461 TCAATTTGCCAAAAGTGTAAAAGAGACACTATGCAAGTGTAGTGCCTATA 510
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
511 ACGCAACAGTTTGAACCTATTGTCATACGCTACCAATAATTCGCGGTAA 560
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
561 ATTTTAAATTCAAATCAATGAAGATGATAAAGCAAATAATGSCATCATTAAC 610
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleaspValLeuA 205
611 TTTCAAGGGGACAGCAATCTGTTAGTAGCAGTACATTACAAGAACAATGG 660
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyRThr 221
661 AATTACACCTGATCTTCGTGGAAATTA...TGGGGAAATAAATTTGAA 707
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyRLeuAs 238
708 GGTGGCGCAATTCGAGAAAGATTTCGAGTCAATTCGTGATTATTATTTAAA 757
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
758 TTAATGGCTATGCCAAGCACAATTTACTAAACCGGATGTTTCAGCTAAATG 807
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
808 ATGAAAACAAACAAAGTTAATGTACCAATTGATTAATGAAGGTTTACAG 857
272 TyRArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyRThrGlnAl 288
858 TATGACCTTCGTAGTGCAGCATTATAGGTAATCTGGGAGGTATGTCGTCG 907
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
908 CGAGCTTGAACCTTTACTTTTCAGCATTTACATTTAAATGATACITTCGCC 957
304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
958 GTAGTGATATTGCAGATTAGAAAATGCAATTTAAAGCAAAACCTTGGAGAA 1007
321 AspGlyTyRtyrTyRAlaGlnIleArgProValThrArgIleAsnAspGl 337
1008 CCGGTTTACGTTAGCGCAACGGTAAATTCAGTACTACTGATTTGTATGATGC 1057
337 uSerArgThrValAspValGluTyRtyrIleAspProValHisProValT 354
1058 AAATAAAACATTAGCGATVAAACCCCTGTTGTTGATGTCGTGGACAGCTTAA 1107

[illegible]

469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
:
1452 AATGAATTATGTGCACAGAGTCAAATTTGGGTTATACCGAGCCCTATTTTAC 1501
:
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
:
1502 TAAAGATGGTGTA.....AGTCTGGTGGAAATGTTTTCTTTGAAAACT 1545
:
502 TyrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
:
1546 ACGATAACTCTAAAAGTGATACATCCTCTAACAATATAAGCGTAGCACTTAC 1595
:
516 GlyGlySerLeuSerTyrGlyTyrProIleaspGluAsnGlnArgIleSe 532
:
1596 GGAAGTAATGTACTTTAGGTTTCCTGTAATAGAAATAACTCCTTATTA 1645
:
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
:
1646 TGTAGGATAGGTCATCATCTATAATAAATAGTAACTTTGCCCTAGAAT 1695
:
544 IyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetaLaAspGly 560
:
1696 ATAACCGTAATTTATATATATCAATCAATGAAA..... 1727
:
561 GlyLysIleGlnValAspAsnAsnGlyIleProaspPheLysHisAspTy 577
:
1728TTTTAAGGTAATGGCATTAACAAACAATGACTTGTGATTT 1765
:
577 rThrThrTyrAsnAlaIleLeuGlyTyrPAsnTyrSerSerLeuAsparg 594
:
1766 TTCT.....TTTGGTTGGAACTATAACAGCCTTAATAGAG 1800
:
594 rovalPheProthrGlnGlyMetSerHisSerValasp.....LeuThr 608
:
1801 GCTATTTCCCRACTAAGGGGTTAAGCAAGTCTGTGGACAGATTACT 1850
:
1851 ATTCCAGGTTCTGTGAACAAATACTACAAACTAAGTCAGATGTACAGGG 1900
:
625 eIfyArgProPhe.....IleLysLysServall 635
:
1901 TTCTTACCCCTATAGACAGATCACCTCTGGTGTGTATCTGCAAAAGCAT 1950
:
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
:
1951 CTGCAAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACC GTTCTAT 1997
:
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
:
1998 CAAACTTATACAGCGGGTGCATCGGTTCATACGTACGTGGTTTGC GTTATGG 2047
:
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlyGlnG 684
:
2048 TAGTATTGGACCTAACGCA.....ATTATGCCGATATGGTAATG 2088
:
684 InThrThrLeuGly.....GluValValGlyGly 693
:
2089 GTAGTGGTACTGGTACTTTTAAAGAAATAAGTCTGTGATGTGATTGGTGT 2138
:
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
:
2139 AATGCANTCGCTACAGTAGCGGACAGAGTTAATGTGGCCA ACTCCATTTGT 2188
:
710 sGlyAsnTrpIle...AspGlnValArgProValIlePheIleGluGly 726
:
2189 GAGCGATAAGAGCAAATAACGGTCCGAACCTCCTTATTTGTGTGATGCGG 2238
:
726 IyGlnValPheaspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
:
2239 CAAGTGTGTGAATCACTAATATGGAATAACAGATAAAATGATTAGAGAGC 2288
:
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758

```
2289 GATGATATAAAA..... 2300
758 aAsnArgProLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
2301 .AGATTCGCTGATTATGGCAAAATCAAGCCGATTTCGGCGCTCTACAGGTG 2349
775 alGlyAlaThrTrpTyrThrProLeuGlyProLeuSerIleSerTyrAla 791
2350 TCGGATTCGAATGCAATCTCCATTGGCCATTGGTATTCTCTATGACC 2399
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG1 808
2400 AAACCAATTAATAATGAAATGATGATGTCGACAGTTCGAACAGTTCCAATTAG 2449
808 nileGlySerValPhe 813
2450 TATTTGGAGGTTCTTTC 2465
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-942-046-1
seq_documentation_block:
; Sequence 1, Application US/08942046
; Patent No. 6284954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS-jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2949 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 75..2465
US-08-942-046-1
```

```
alignment_scores:
Quality: 1052.50      Length: 822
Ratio: 1.956          Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630
alignment_block:
US-09-701-711-2 x US-08-942-046-1 ..
Align seg 1/1 to: US-08-942-046-1 from: 1 to: 2949
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
114 ACAACGACTGCTGTTGGCCACCTTTTGTGCAAAAGATATTCGTGTGA 163
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
164 TGGTGTTCAGGTGCTAGAACACAANAATCCGAGCAAGTTTACTGTTC 213
56 rgLeuGlyGlnValValserGluAsnGlnLeuAlaAspGlyValLysAla 72
214 GTGCCGGTCAAGCGTGTGACTGACAATGATGTGGCTAATATTTGTCGCT 263
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG1 89
264 TTATTCGTAAGTGTGCTGATTCGATGTGAAAGCG...CATCAAGAAG 310
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
311 CGATGCTGTTGTTAGCGTTGGCTAAATTCGATCATTTTCAGATGTTA 360
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
361 AAATCAAAAGGTAACTGCTTTATCCCACTGAAGCACTTAAACAAACCTA 410
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG1 138
411 GATGCTAACGGGTTAAAGTTGGCGATGTTTAAATTCGAGAAAATTTAA 460
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
461 TGAATTTGCCAAAAGTGTAAAGAGCACATATGCAAGTGTAGTCGCTATA 510
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
511 ACGCAACAGTTGAACCTATTCTCAATACGCTACCAATAATATCGCGCTGA 560
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
561 ATTTTAAATTCAAATCAATGAAGATGATAAAGCAAAATTTGGCATCTAAC 610
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
611 TTCAAGGGGAACCAATCTGTTAGCAGTACATTTACAAGAAACAATGG 660
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
661 AATTACAACCTGATCTCTGGTGGAAATTA...TGGGGAATAAATTTGAA 707
222 GlnGluLysLeuValThrSerLeuGlnAsnLeuArgAlaLysTyrLeuAs 238
708 GGTGCGCAATTCGAGAAAGATTTTCAGTCAATTCGTGATTATTATTTAA 757
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
758 TAATGGCTATGCCAAAGCACAAATTTACTAAACGGATGTTTCAGCTAAATG 807
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
808 ATGAAAAACAAAAGTTAATGTAACCATTTGATGTAATGAAGGTTTACAG 857
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
858 TATGACCTTCGTAGTCGACGCAATTATAGGTAATCTGGGAGGTATGCTGCG 907
```

```
288 agluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
908 CGAGCTTGAACCTTTACTTTCAGCATTTAAATGATCTTCCGCC 957
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 InAlaMetLeuGluGlnThrThrAsnAsnLleSerThrLysPheGlyAsp 320
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
958 GTAGTGATATGCAGATGAGAAAATGCAATTAAGCAAAACTTGAGAA 1007
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 AspGlyTyrTyrAlaGlnLleArgProValThrArgLleAsnAspG1 337
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1008 CGCGGTACGGTAGCCACGCGTAATTCAGTACCTGATTTTGATGATGC 1057
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 uSerArgThrValAspValclutTyrTyrLleAspProValHisProValT 354
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1058 AAATAAACATAGCGATAACCTTCTTGTGATGCTGGACGACGTTAA 1107
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 yrValArgArgLleAsnPhetheGlyAsnPhelysthrGlnaspGluVal 370
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1108 CTGTTCCCACTTCGCTTGAAGGAATACCGTTTCTCTGATAGCACT 1157
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1158 TTACGTCAGGAATGCGCCCAAGAAGAACTTGTATAATTCACAAT 1207
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1208 AGTTGAGTTAGAAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1253
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1254 ..ACAGTGCAGAACCGAATGTATCTCAATGAGTAGTAAATGAAAGTG 1301
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 AspValAsnPheValValGluGlnProSerGlySerThrIleAl 435
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1302 GATGCGTATATAAGTCAAGAAAGTAAACACGCGGTAGTATCAACTTGG 1351
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1352 TATGTTGACGTACAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1401
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1402 AGATAAATTTCTTGGACAGCGCGGAGTAGTATAGTACGTGACGAAA 1451
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1452 AATGATTATGTGTACGAGTCAATTTGGGTTATACCGAGCCCTATTTTAC 1501
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrArgLysThrLysT 502
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1502 TAAAGATGGTGA.....AGTCTTGGTGAAATGTTTCTTTGAAAACT 1545
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 yTAspAsnLys.....AsnLleSerAsnTyrValLeuAspSerTyr 515
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1546 ACGATAACTCTAAAGTATACATCTCTAATAAGCTAGCGTACGACTTAC 1595
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 GlyGlySerLeuSerTyrGlyTyrProLleAspGluAsnGlnArgLese 532
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1596 GGAAGTAATGTACTTTAGGTTTCCCTGTAATGAAATAAATCACTATT 1645
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
532 rPheGlyLeu.....AsuAlaAspAsnThrLysLeuHisG 544
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1646 TGTAGGATTAGCTACACTATAATAAATAATAGTAACCTTGTCTAGAA 1695
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
544 lyGlyArgPheMetGlyLleSerAsnValLysGlnLeuMetaAspGly 560
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1696 ATAACCGTAATTTATATATCAATCAATCAAGAAA..... 1727
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
561 GlyLysLleGlnValAspAsnAsnGlyLleProAspPheLysHisAspTy 577
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1728 .....TTTAAAGGTAATGGCAATTAACAAACAAATGACTTTGATT 1765
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-08-433-522A-5

seq_documentation_block:

; Sequence 5, Application US/08433522A

; Patent No. 6013514

; GENERAL INFORMATION:

; APPLICANT: CHONG, Pele

; APPLICANT: THOMAS, Wayne

; APPLICANT: YANG, Yan Ping

```

; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..2724
; US-08-433-522A-5

```

alignment_scores:

```

Quality: 1052.50      Length: 822
Ratio: 1.956          Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

```

alignment_block:

```
US-09-701-711-2 x US-08-433-522A-5 ..
```

```
Align seg 1/1 to: US-08-433-522A-5 from: 1 to: 2950
```

```

23 SerThrHisAlaGlnAlaAspPheMetAlaAsnAspIleThrIleTh 39
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
373 ACAACGACTGTGTTGGCCGACCTTTTGGCAAGATATTTCGTGGA 422
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
423 TGGTGTTCAGGTGACTAGACACACAAATCCGACGAGTTTACCTGTTC 472
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
56 rgLeuGlyGlnValSerGluAsnGlnLeuAlaAspGlyVallysAla 72
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
473 GTGCCGGTCAGCGTGTGCTAGCAATGATGTGCTTAATATTGTCGCTCT 522
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
523 TTATTCGTAAGTGGTCGATTCGATGATGATGGAAGCG...CATCAAGAAG 569
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
89 yArgile...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
570 CGATGTGCTGTGTAGCGTTGTGCTTAATTCGATCATTTTCAGATGTTA 619
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGlyLeu 121
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

620 AAATCAAAAGGTAACTCTGTATTATTCACGACTTAAACAAAACTTA 669
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGI 138
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
670 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTA 719
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
720 TGAATTTGCCAAAAGGTAAAGAGACACTATGCAAGTGTAGTGCCTATA 769
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
770 ACGCAACAGTTGAACCTATTGTCAATACGCTACCAATAATTCGCTGAA 819
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
820 ATTNTAATTCAAATCAATCAATGAAGATGATAAGCAAAATTCGATCATA 869
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
870 TTTCAGGGGGAACGAATCTGTTAGTAGCAGTACATTACAAGACAATGG 919
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
920 AATTACACCTGATTTCTGGTGGAAATTA...TGGGAAATAAATTTGAA 966
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
967 GTGCGCAATTCGAGAAAGATTTGCAGTCAATTCGGGATTTATTATAA 1016
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1017 TAATGCTATGCCAAGACAAATTAATAAAGGATGTTTCAGCTAAATG 1066
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1067 ATGAAAAACAAAAAGTTAATGTAACCATTTGATGATAAATGAAGTTTAC 1116
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1117 TATGACCTTCGTAGTGCACGATTAAGTAATCTGGGAGGTATGCTGTC 1166
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1167 CGAGCTTGACCTTTACTTTCAGCATTAATTAATGATACTTTCGCC 1216
304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1217 GTAGTGATATTGCAGATGTAGAAAATGCAATTAAGCAAACTTGGAGA 1266
321 AspGlyTyrTyrThrAlaGlnIleArgProValThrArgIleAsnAspGI 337
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1267 CCGGTTACGGTAGCGCAACGGTAAATTCAGTACCTGATTTGATGATGC 1316
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProVal 354
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1317 AAATAAACAATTAGCATTAACCTTTGTTGATGCTGGACGAGCTTAA 1366
354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1367 CTGTTCCGCAACTTCGCTTTGAAGGAATACCGTTTCTGCTGATAGCA 1416
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1417 TTACGTCAGGAATGCGCAACAAGAACTTGGTATAAATTCACAAT 1466
387 stIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHis 404
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1467 AGTTGAGTTAGGAAAAATTCGCTTAGTACGACAGGTTTCTCGAA... 1512
404 aThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1513 ..ACAGTCGAAAAACCAATTTGATCCCTATCAATGTTAGTAGTAATGATGA 1560

```

```
419 AspValAsnPhValValGluGlnProSerGlySerSerThrIleAl 435
||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1561 GATGTCGTATATAAGTCAAGAAGCGTAAACGCGGTACTATCAACTTGG 1610
435 aaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
|||:||||| :|||:||||| :|||:||||| :|||:||||| :|||:|||||
1611 TATTGGTTACGGTACAGAGAGTGTATTAGTTATCAACAAGAGTGTAAAC 1660
452 InAsnAsnPhMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1661 AAGATAATTTCTTGGGAACAGCGCGGAGTAAGTATAGCTGGTACGAAA 1710
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1711 AATGATTATGTACGAGGTCAATTTGGTTATACCGAGCCCTATTTTAC 1760
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1761 TAAAGATGGTGTAT.....AGTCTGGTGGAAATGTTTCTTTGAAAACT 1804
502 YrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1805 ACGATAACTCTAAAGTACATACCTCTACTATTAACGCTACGACTTAC 1854
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1855 GGAAGTAAATGTACTTATAGTGTCTCCCTGTAAATGAAATAACTCTCTATTA 1904
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1905 TGTAGGATTAGCTATACCTATATAATAAATTAGTAACCTTCTCTAGAA 1954
544 lyGlyArgPheMetGlyLysSerAsnValLysGlnLeuMetAlaAspGly 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1955 ATAACCGTAATTTATATTAATCAATCAATGAAA.....1986
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1987 .....TTTAAAGGTAATGGCAATTAATAAACAATGACTTTGATT 2024
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2025 TTCT.....TTTGGTGGAACTATAACAGCTTAAATAGAG 2059
594 roValPheProThrGlnGlyMetSerHisSerValasp.....Leuthr 608
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2060 GCTATTTCCCAACTAAAGGGGTAAAGCAAGTCTTGTGGACGAGTTACT 2109
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnII 625
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2110 ATTCAGGTTCTGATAACAATACTACAAACTAAAGTCAGATGTACAGGG 2159
625 eTyrArgProphe.....IleLysLysSerValL 635
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2160 TTTCTACCCATTACAGAGATCACTCTGGTGTATCTGCAAAACAT 2209
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2210 CTGCAGGATATGCAAAAT...GGTTTGGAAACAAGCGTTTACCCTTCTAT 2256
651 GluAsnPhetYrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2257 CAAACTTATACAGCGGTGCGATCGGTTCATTAGCTGGTTTGGTTATGG 2306
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2307 TAGTATGGACCTTAACCA.....ATTATGCCCAATATAGTATG 2347
684 InThrThrLeuGly.....GluValValGlyGly 693
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2348 GTAGTGCTAGTGGTACTTTTAAAGAAGATAAGTTCTGTATGTGTTGGT 2397
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-135-166-5

seq documentation block:

; Sequence 5, Application US/09135166

; Patent No. 6083743

; GENERAL INFORMATION:

; APPLICANT: CHONG, Pele

; APPLICANT: THOMAS, Wayne

; APPLICANT: YANG, Yan Ping

; APPLICANT: LOOMORE, Sheena

; APPLICANT: SIA, Dwo Yuan Charles

; APPLICANT: KLEIN, Michel

; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN

; NUMBER OF SEQUENCES: 55

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6TH Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/135,166

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/433,522

; FILING DATE: 12-SEP-1995

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: STEWART, Michael I

; REGISTRATION NUMBER: 24,973

; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163


```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..2724
US-09-135-166-5

```

```

alignment_scores:
  Quality: 1052.50      Length: 822
  Ratio: 1.956          Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

```

```
alignment_block:

```

```
US-09-701-711-2 x US-09-135-166-5 ..
```

```
Align seg 1/1 to: US-09-135-166-5 from: 1 to: 2950
```

```

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
373 ACAACGACTGTGTTGCCGACACTTTTGGCAAAAGATATTCGTGTGGA 422
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
423 TGGTGTCAAGGTGACTTGAACAACAAATCCGACGAACTTACCTGTTC 472
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
473 GTGCCGTCAGCGTGTGACTGACATGATGGCTAATATTGTCGCTCT 522
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluI 89
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
523 TTATTTCGTAAGTGTGATTCGATGATGTAAGCG...CATCAAGAAG 569
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
570 CGATGTGCTGTTGTTAGCGTGTGGCTAAATCGATCATTTTCAGATGTTA 619
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
105 snPheGlyAsnArgLeuIleProLysGluGlyLeuGlnGlyLeu 121
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
620 AAATCAAGGTAACCTCTGTTATTCCTCCAGCAAGCACTTAACAAACTTA 669
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
670 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTA 719
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrA 155
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
720 TCAATTTGCCAAAGTGTAAAAGAGACACTATCCAGTGTAGTTCGCTATA 769
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
770 ACGCAACACTTGAACCTATTGTCAATACGCTACCAAAATAATCGCGCTGA 819
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
172 LeuAspMetThrPheAlaGlyLysProAlaArgValValAspIleAs 188
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
820 ATTTTAAATTCAAATCAATGAAGATGATAAGCAAAATTCGCATCATTA 869
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeu 205
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
870 TTTCAAGGGGACGAATCTGTTAGTACGAGTACATTACAAGCAACAATGG 919
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
920 AATTACAACCTGATCTTGGTGGAAATTA...TGGGGAATAAATTTGAA 966
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

967 GGTGCGCAATTCGAGAAAGATTTCAGTCAATTCGTGATTATTATTTAAA 1016
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1017 TAATGGCTATGCCAAACGACAAATTAATAAACGGATGTTTCAGCTAAATG 1066
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1067 ATGAAAAAACAAAAGTTAATGTAAACCATTTGATGTAATGAAGGTTTACAG 1116
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1117 TATGACCTTCGTAGTCACGCATTATAGGTAATCTGGGAGGTATGCTGCG 1166
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1167 CGAGCTTGAACCTTTACTTTTCAGCATTAATTAATGATACTTTTCGCC 1216
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
304 lnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1217 GTAGTGTATTCGAGATGTAGAAAAATGCAATTAAGCAAACTTGGAGAA 1266
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG 337
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1267 CGCGGTTACGGTAGCGCAACGGTAAATTCAGTACCTGATTTGATGATGC 1316
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1317 AAATAAACATTTAGCGATTAACCCCTGTTGTGATGCTGGACGACGTTTAA 1366
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
354 yrValArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1367 CTGTTCCGCCAACTTCGCTTTGAAGGAAATACCGCTTTCGTGATAGCACT 1416
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
371 LeuArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLys 387
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1417 TTACGTCAGGAAATGCCCAACGAAGAACTTGGTATATAATTCACAATT 1466
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1467 AGTTAGTTAGGAAAAAATTCGCTTAGCTACAGGTTCTTCGAA... 1512
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1513 ..ACAGTCGAAACCGCAATTTGATCCTCATCAATGGTAGTAATGATGAAGTG 1560
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1561 GATGTCGTATATAAGTCAAAAGAACGTAACACGGGTAGTATCAACTTTGG 1610
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1611 TATTGTTACGGTACAGAGATGGTATTAGTTATCAAGCAAGTGTTAAC 1660
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
452 lnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1661 AAGATAAATTTCTGGCAACAGGGCGGCAGTAAGTATAGCTGGTACGAAA 1710
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1711 AATGATTATTCGTACGAGTGTCAATTTGGGTTATATACCGAGCCCTATT 1760
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1761 TAAAGATGGTGA.....AGTCTTGGTGGAAATGTTTCTTTGAAAAC 1804
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1805 ACGATAACTCTAAAGATGATACATCTCTAATAAGCGTACGACTTAC 1854
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1855 GGAAGTAATGTTACTTTAGTTCCTGTAATGAAATAAATAACTCCTCTATTA 1904
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
||||| : : : : : |||||
1905 TGTAGATTAGTGCATACCTATAATAAAATAGTAACCTTCTCTAGAAT 1954
||||| : : : : : |||||
544 lyGlyArgPheMetGlySerAsnValLysGlnLeuMetAlaAspGly 560
||||| : : : : : |||||
1955 ATAACCGTAATTATATATCAATCAATGAAA..... 1986
||||| : : : : : |||||
561 GlyLysIleGlnValAspAsnGlyIleProAspPheLysHisAspTy 577
||||| : : : : : |||||
1987 .....TTAAAGGTAATGGCAATAAACAAAGACTTTGATT 2024
||||| : : : : : |||||
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
||||| : : : : : |||||
2025 TTCT.....TTTGGTGGAACTATAACAGCCTTAATAGAG 2059
||||| : : : : : |||||
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
||||| : : : : : |||||
2060 GCTATTTCCTCAACTAAAGGGTTAAAGCAAGTCTTGGTGGAGCTTACT 2109
||||| : : : : : |||||
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
||||| : : : : : |||||
2110 ATTCCAGGTTCTGATAACAATACTACAACTAAAGTGCAGATGTACAGGG 2159
||||| : : : : : |||||
625 eTyrArgProPhe.....IleLysLysSerValL 635
||||| : : : : : |||||
2160 TTCTACCATCTAGACAGATCACTCTGGGTTGTCTATCTGCAAAAGCAT 2209
||||| : : : : : |||||
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
||||| : : : : : |||||
2210 CTGCAGGATATGCAAA...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2256
||||| : : : : : |||||
651 GluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSe 667
||||| : : : : : |||||
2257 CAACCTTATACAGCGGGTGCATCGCTTCAATTCGTTTTCGCTATGG 2306
||||| : : : : : |||||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
||||| : : : : : |||||
2307 TAGATTGGACCTAAGCA.....ATTATGCCGAATATGTTAATG 2347
||||| : : : : : |||||
684 lnThrThrLeuGly.....GluValValGlyGly 693
||||| : : : : : |||||
2348 GTAGTGGTACTGGTACTTTTAAAGAGATAAGTCTCATGTGATGGTGGT 2397
||||| : : : : : |||||
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
||||| : : : : : |||||
2398 ATGCANTCGCTACAGCTAGCGCAGAGTTAATGTGCCAACTCCATTGT 2447
||||| : : : : : |||||
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
||||| : : : : : |||||
2448 GAGCGATAAGACCAAAATACGTCGCAACCTCCTTATTGTTGATGCGG 2497
||||| : : : : : |||||
726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
||||| : : : : : |||||
2498 CAAGTGTGTTGGAACTAATGAATCAGATAAAATGATGATGAGAGACC 2547
||||| : : : : : |||||
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaA 758
||||| : : : : : |||||
2548 GATGTATTAA..... 2559
||||| : : : : : |||||
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
||||| : : : : : |||||
2560 .AGATTGCTGTATTATGGCAATCAAGCGGTATTCGCGCTCTTACAGGTG 2608
||||| : : : : : |||||
775 aGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
||||| : : : : : |||||
2609 TCGGATTCCAATGGCAATCTCTATTGGCCCATTTGTTATCTCTTATGCC 2658
||||| : : : : : |||||
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
||||| : : : : : |||||
2659 AAACCAATTAATAAATATGAAATGATGATGTGCAACAGTTCCAATTTAG 2708

```

```

808 nileGlySerValPhe 813
:|||||: |||
2709 TATTGGAGGTTCTTTC 2724

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-942-046-5

seq_documentation_block:
; Sequence 5, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2950 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 334..2724
; US-08-942-046-5

alignment_scores:
Quality: 1052.50 Length: 822
Ratio: 1.956 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-08-942-046-5 ..
Align seg 1/1 to: US-08-942-046-5 from: 1 to: 2950
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
:|||||: ||||| :|||: |||||
373 ACAACCACTGTGTTGCCGACCTTTTGGCAAAAGATATTCTGTGGA 422
:|||||: ||||| :|||: |||||
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProphea 56

```



```
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
|||||
2210 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2256
|||||
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
|||||
2257 CAACCTATACAGCGGGTGGATCGGTTTCATAGCTGGTTGCTTATGG 2306
|||||
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgGlyGlnG 684
|||||
2307 TAGTATTCGACCTAAGCA.....ATTATGCCGAATATGTAATG 2347
|||||
684 lnThrThrLeuGly.....GluValValGlyGly 693
|||||
2348 GTAGTGGTACTGGTACTTTTAAAGATAAGTTCTGATGTGATGGTGGT 2397
|||||
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
|||||
2398 AATGCAATCGTACAGTAGCGCAGAGTTAATTGTGCCAACTCCATTGT 2447
|||||
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGly 726
|||||
2448 GAGCGATAAGACCAATAACCGTCCGAACCTCCTTATTGTTGATGCGG 2497
|||||
726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
|||||
2498 CAAGTGTGTGGAACTACTAAATGGAATCAATAAATGGATTAGAGAGC 2547
|||||
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
|||||
2548 GATGTATTAAAA..... 2559
|||||
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
|||||
2560 .AGATTGCTGCTATTATGCAAAATCAAGCGGTATTGCGGCCCTCAGAGTG 2608
|||||
775 alGlyAlaThrTrpThrProIleGlyProLeuSerIleSerTyrAla 791
|||||
2609 TCGGATTCATGCAATCTCTATTGGCCATGGTATTCTCTATGCC 2658
|||||
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG 808
|||||
2659 AAACCAATTAATAATGAAATGATGATGTCGAACAGTTCCAATTTAG 2708
|||||
808 nIleGlySerValPhe 813
|||||
2709 TATTGGAGGTTCTTTC 2724
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-433-522A-3

```
seq_documentation_block:
; Sequence 3, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 374..2764
; US-08-433-522A-3
```

```
alignment_scores:
Quality: 1052.50 Length: 822
Ratio: 1.956 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-08-433-522A-3 ..
```

```
Align seg 1/1 to: US-08-433-522A-3 from: 1 to: 2984

23 SerThrHisAlaGlnAlaAspPheMetAlaAsnAspIleThrIleTh 39
413 ACAAGGACTGTGTTGCGCGACCTTTTGTGGCAAAAGATATTTCGTGGA 462
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
463 TGGTGTTCAGGTCGACTTAGAACACAAATCCGACAAAGTTTACCTGTC 512
56 rGlyLeuGlyValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
513 GTGCGGTCAGCGTGTGACTGACAAATGATGTGGCTAATATTTCGCGCTC 562
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG 89
563 TTATTCGTAAGTGTGATTCGATGATGAAAGCG...CATCAAGAAGG 609
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
610 CGATGTGCTTGTGTTGCGGTTGGCTAAATTCGATCATTTTCAGATGTTA 659
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
660 AAATCAAAAGGTAACCTCTGTTATTCCTCCACTGAAGCAGCTTAAACAAA 709
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
710 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATAAA 759
138 nMetIleGluThrGluLeuThrAspGlnTyrIleSerGlnGlyTyrTyrA 155
760 TGAATTTGCCAAAAAGTGTAAAGACACTATGCCAAGTGTAGTTCGTATA 809
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
810 ACGCAACAGTTGACCTATTGTCAATAGCTACCAATAAATCGCGTGA 859
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
```


758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
::: ||| ::::::::::::::: ||| |||::: |||
2600 .AGATTGCTGATTATGCAAAATCAAGCGGATTTCGGCGCTCTACAGGTG 2648
775 alGlyAlaThrTyrProIleGlyProLeuSerIleSerTyrAla 791
||| ||| ::::::::::::::: ||| |||::: |||
2649 TCGATTCCATGCAATCTCTATTGGCCATTGGTATTCTCTATGCC 2698
792 LysProLeuAsnLysLysGlnAspGlnThrAspThrValGlnPheG1 808
||| ||| ::::::::::::::: ||| |||::: |||
2699 AAACAATTAATAATGAATAATGATGTCGAACAGTTCCTCAATTAG 2748
808 nileGlySerValPhe 813
: ||| |||
2749 TATTGGAGGTCTTTC 2764

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-135-166-3

seq_documentation_block:
; Sequence 3, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2984 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 374..2764
; US-09-135-166-3

alignment_scores: Quality: 1052.50 Length: 822

Ratio: 1.956 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630
alignment_block:
US-09-701-711-2 x US-09-135-166-3 ..
Align seg 1/1 to: US-09-135-166-3 from: 1 to: 2984
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
::: ||| ::::::::::::::: ||| |||::: |||
413 ACAACGACTGTGTTTGGCGCACCTTTGTGGCAAAAGATATTCGTGGGA 462
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
||| ||| ::::::::::::::: ||| |||
463 TGGTGTTCAGGTGACTTAGAACACAAATCCGAGCAAGTTTACCTGTTC 512
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
||| ||| ::::::::::::::: ||| |||
513 GTGCGGTCAGCGTGTGACTGACATGATGTGGCTAATATTATTCGCGCTCT 562
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG1 89
||| ||| ::::::::::::::: ||| |||
563 TTATTGTAAGTGTGCTGATTCGATGTGAAGCG...CATCAAGAAGG 609
89 yArgile...IleTyrGlnValThrGluArgProLeuIleAlaGluLea 105
||| ||| ::::::::::::::: ||| |||
610 CGATGTGCTTGTGTTAGCGTTGTGGCTAAATCGATCATTTTCAGATGTTA 659
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
::: ||| ::::::::::::::: ||| |||
660 AAATCAAGGTAAGTCTGTTATTCCTGAAAGCACTTAAACAAACTTA 709
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG1 138
||| ||| ::::::::::::::: ||| |||
710 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAATATAA 759
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
::: ||| ::::::::::::::: ||| |||
760 TGAATTTGCCAAAAGTGTAAGAGACACTATGCAAGTGTAGTGTGCTATA 809
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
||| ||| ::::::::::::::: ||| |||
810 ACGCAACAGTTGAACCTATTGTCAATAGCTACCAAAATATCGCGCTGAA 859
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
::: ||| ::::::::::::::: ||| |||
860 ATTTAATTCAAAATCAATGAAGATGATAAGCAAAATTCGCATCATTAAC 909
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
::: ||| ::::::::::::::: ||| |||
910 TTTCAAGGGGAACGAATCTGTTAGTAGCAGTACATTACAAGCAAAATGG 959
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
::: ||| ::::::::::::::: ||| |||
960 AATTACAACCTGATCTTGTGGTAATTA...TGGGAAATAAATTTGAA 1006
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
::: ||| ::::::::::::::: ||| |||
1007 GGTGCGCAATTCGAGAAAGATTTGCAGTCAATTCGTGATTATTATAA 1056
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
||| ||| ::::::::::::::: ||| |||
1057 TAATGGCTATGCCAAAGCAAAATTAATAAACGGATGTTACGCTAATG 1106
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
::: ||| ::::::::::::::: ||| |||
1107 ATGAAAAACAAGTAAATGTAACCATTTGATGTAATAATGAAGTTTACAG 1156
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
||| ||| ::::::::::::::: ||| |||
1157 TATGACCTTCGTAGTCACGCAATTAATAGTAATCTGGGAGGTATGCTGTC 1206
288 aGluLeuGluAlaLeuLys...PheLysAlaGluGlyPheSerG 304

```
||||| 1207 CGAGCTTGAACCTTTACTTTTCAGCATTACATTTAAATGATACCTTTCGCC 1256
||||| 304 lnAlaMetLeuGluGlnThrThrasnAsnIleSerThrLysPheGlyAsp 320
||||| 1257 GTAGTGATATTGCAGATGTAGAAATGCAATTAAGCAAACTTTGGAGAA 1306
||||| 321 AspGlyTyrTyrTrpAlaGlnIleArgProValThrArgIleAsnAspG1 337
||||| 1307 CGCGGTTACGTACGCCAACGGTAAATTCAGTACCTGATTTTGATGATGC 1356
||||| 337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
||||| 1357 AAATAAAACATATTAGCATACACCTTGTTGTTGATGCTGCGACGAGCTTAA 1406
||||| 354 yrValArgArgIleAsnPhethrGlyAsnPhelysThrGlnAspGluVal 370
||||| 1407 CTGTTCCCAACTTCGCTTTGAAGAAATACCGTTTCTGCTGATGACACT 1456
||||| 371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
||||| 1457 TTAGCTCAGGAATGCCCAAGAGAGAACTTGGTATATATTCACAATT 1506
||||| 387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
||||| 1507 AGTTGAGTTAGGAAAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1552
||||| 404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
||||| 1553 ..ACAGTCGAAACCGGAATTCATCTCAATGATGATGATGATGATGATG 1600
||||| 419 AspValAsnPhelValValGluGlnProSerGlySerThrIleAl 435
||||| 1601 GATGCTGATATAAAGTCAAGAACGTAACACGGGTAGTATCAACTTTGG 1650
||||| 435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
||||| 1651 TATTGGTTAGCGTACAGAGAGTGTATTAGTTATCAAGCAAGTGTAAAC 1700
||||| 452 lnAsnAsnPhelMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
||||| 1701 AAGATAATTTCTGGGAACAGGGCGCGCATAGTATAGCTGGTACGAAA 1750
||||| 469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
||||| 1751 AATGATTATGTCAGAGTGTCAATTTGGGTATACCGAGCCCTATTTTAC 1800
||||| 485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
||||| 1801 TAAAGATGGTGA.....AGTCTGTGGGAAATGTTTCTTTGAAAAC 1844
||||| 502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
||||| 1845 ACAGATACTCTAAAGTGTATACATCTCTAACTATAAGCGTACGACTTAC 1894
||||| 516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
||||| 1895 GGAAGTAATGTACTTTAGGTTCCCTGTAAATGAAATAACCTCTATTA 1944
||||| 532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
||||| 1945 TGTAGGATTAGGTCATACCTATAATAAAATAGTAACCTTTGCTCTAGAA 1994
||||| 544 lyGlyArgPheMetGlyLysSerAsnValLysGlnLeuMetAlaAspGly 560
||||| 1995 ATACCGTAATTTATATATTCAATCAATGAAA..... 2026
||||| 561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
||||| 2027 .....TTTAAAGGTAATGGCATTAACAACAAATGACITTTGATTT 2064
||||| 577 rThrThrTyrAsnAlaIleLeuGlyTyrPAsnTyrSerSerLeuAspArgp 594
|||||
```

```
2065 TTCT.....TTTGGTTGGAACATAACAGCCTTAATAGAG 2099
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
||||| 2100 GCTATTTCCCAACTAAAGGGTTAAAGCAAGTCTTGGTGGAGAGTTACT 2149
||||| 609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
||||| 2150 ATTCAGGTTCTGATATAACAATACTACAACAATAAGTGCAGATGTACAGG 2199
||||| 625 eTyrArgProPhe.....IleLysLysSerValL 635
||||| 2200 TTTTACCACTATAGACAGAGATCACTCTGGGTTGTATCTGCAAAAGAG 2249
||||| 635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
||||| 2250 CTGCAGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTTCAT 2296
||||| 651 GluAsnPhetyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
||||| 2297 CAAACTATATACAGCGGTGGCATCGTTCAATACGTGTTTTCGTTATGG 2346
||||| 667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
||||| 2347 TAGTATTGGACCTAACGCA.....ATTATCCGAATATGTAATG 2387
||||| 684 lnThrThrLeuGly.....GluValValGlyGly 693
||||| 2388 GTAGTGTGACTGGTACTTTTAAAGAAATAAGTTCTGTGATGTTGGTGGT 2437
||||| 694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
||||| 2438 AATCAATCGCTACAGTACGAGCGAGAGTTAATGTGCCAACTCCATTGT 2487
||||| 710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
||||| 2488 GAGCGATAAGAGCCAAATAACGTCGCCAACCTCTTATTGTTGATGCGG 2537
||||| 726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
||||| 2538 CAAGTGTTTGGATACTAAATGGAATAACAGATAAAATGGATAGAGAGC 2587
||||| 742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
||||| 2588 GATGATTAATAA..... 2599
||||| 758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
||||| 2600 .AGATTGCTGATTATGCAAAATCAAGCCGATTTCGCCCTCTACAGGTG 2648
||||| 775 alGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
||||| 2649 TCGGATTCCAATGGCAATCTCTATNTGGCCATTTGGTATTCTTTATGCC 2698
||||| 792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheG1 808
||||| 2699 AAACCAATTAATAAATATGAAATGATGATGTCGCAACAGTTCCCAATTAG 2748
||||| 808 nIleGlySerValPhe 813
||||| 2749 TATTGGAGGTTCTTTC 2764
seq_name: /cnp2_6/ptodata/2/ina/6B_COMB.seq:us-08-942-046-3
seq_documentation_block:
; Sequence 3, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
```

[illegible][illegible]


```
|||||..... |||... :||| |||... |||
1553 ..ACAGTCGAAACCGAATTTGATCCTATCAATGGTAGTAATGATGAAGTG 1600
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
||||| :|||... |||... :|||... |||... |||
1601 GATGTCGTATATAAGTCAAGAACCAACACACGGGTAGTATCAACTTTGG 1650
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
: ||||... :|||... :|||... :|||... :|||... :|||...
1651 TATTGTTACGGTACAGAGAGTGGTATTAGTTATCAAGCAAGTGTAAAC 1700
452 lnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
|||... |||... |||... |||... :|||... :|||...
1701 AAGATAAATTTCTTGGCAACAGGCGCAGTAAAGTAGCTGTACGAAA 1750
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
: |||... :|||... :|||... :|||... :|||... :|||...
1751 AATGATTATGGTACGAGTGTCAATTTGGGTTATACGAGCCCTATTTTAC 1800
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
| :|||... |||... :|||... :|||... :|||... :|||...
1801 TAAAGATGGTGA.....AGTCTTGGTGAATGTTTCTTTGAAAAC 1844
502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
|||||... :|||... :|||... :|||... :|||... :|||...
1845 ACGATAACTCTAAAGTGATACATCCTCTAACTATAAGCGTACGACTTAC 1894
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
||||| :|||... |||... :|||... :|||... :|||... :|||...
1895 GGAAGTAATGTTACTTTAGGTTTCCCTGTAATGAAATGAAATCCCTATTA 1944
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
||||| :|||... :|||... :|||... :|||... :|||... :|||...
1945 TGTAGATTAGGTCAACATATATAATAATAGTAATTTGCTCTAGAAT 1994
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
: |||... :|||... :|||... :|||... :|||... :|||...
1995 ATAACCGTAATTTATATATATCAATCAATGAAA..... 2026
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
||||| :|||... :|||... :|||... :|||... :|||... :|||...
2027 .....TTTAAAGGTATGCAATTAACAAACAAATGACTTTGATTT 2064
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgP 594
: |||... :|||... :|||... :|||... :|||... :|||...
2065 TTCT.....TTTGGTGGAACTATACAGCCCTTAATAGAG 2099
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
||||| :|||... :|||... :|||... :|||... :|||... :|||...
2100 GCTATTTCCTCACTAAAGGGGTTAAAGCAAGTCTTGGTGACGAGTTACT 2149
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnII 625
: |||... :|||... :|||... :|||... :|||... :|||...
2150 ATCCAGGTCTGTAACAAATACTACAACTAAAGTGACAGATGACAGGG 2199
625 eTyrArgProPhe.....IleLysLysSerValI 635
: |||... :|||... :|||... :|||... :|||... :|||...
2200 TTCTACCCATTAGACAGATACACCTCTGGTGTGTATCTGCAAAAGCAT 2249
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
||||| :|||... :|||... :|||... :|||... :|||... :|||...
2250 CTGACGATATGCAAA...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2296
651 GluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSe 667
: |||... :|||... :|||... :|||... :|||... :|||...
2297 CAAACTTATACAGCGGGTGCATCGGTTTATACGTTGGTTTCTTATGG 2346
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
||||| :|||... :|||... :|||... :|||... :|||... :|||...
2347 TAGATTGGACCTAAGCA.....ATTATGCCGATATGGTAATG 2387
684 lnThrThrLeuGly.....GluValValGlyGly 693
: |||... :|||... :|||... :|||... :|||... :|||...

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:us-08-433-522A-55
seq_documentation_block:
; Sequence 55, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
```

```

; LENGTH: 2987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-433-522A-55

```

```

alignment_scores:
  Quality: 1052.50      Length: 822
  Ratio: 1.956          Gaps: 20
  Percent Similarity: 65.450  Percent Identity: 31.630

```

```

alignment_block:
US-09-701-711-2 x US-08-433-522A-55

```

```

Align seg 1/1 to: US-08-433-522A-55 from: 1 to: 2987

```

```

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
428 ACAACGACTGTGTTGGCGCACCTTTTGGCAAAAGATATTTCGTGTGA 477
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
478 TGGTCTCAAGTGACTTAGAACACAAATCGAGCAAGTTTACCTGTTC 527
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspGlyVallysAla 72
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
528 GTGCGGTCAGCGTGTGACTGACATGATGATGCGGTAAATTTTCGCTCT 577
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
73 LeuTyAlaThrGlyAsnPheSerAspValGlnValTyHisGlnGluGl 89
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
578 TTATTCGTAAGTGGTCGATTCGATGATGTAAGCG...CATCAAGAAGG 624
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
89 yArglle...lleTyrglnValThrGluArgProLeuIleAlaGluIleA 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
625 CGATGTGCTTGTGTTAGCGTTGTGGCTAAATCGATCATTTTCAGATGTTA 674
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGlyLeu 121
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
675 AAATCAAGAGTAACTCTGTATTCCTCCACTGAAGCACTTAACAAACTTA 724
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGl 138
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
725 GATGCTAAACGGGTTTAAAGTTCGCGATGTTTAAATTCGAGAAATAAA 774
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
138 nMetIleGluThrGluLeuThrAsnGlnTyrlleSerGlnGlyTyTyra 155
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
775 TGAATTTGCCAAAGTGTAAAGAGGACCTATGCCAAGTGTAGTTCGTATA 824
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 snThrGluIleThrVallysGlnThrMetLeuAspGlyAsnArgVallys 171
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
825 ACACAACAGTTGAACCTATTGTCAATACGCTACCAATAATCGCGCTGAA 874
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
875 ATTTTAATTCAAATCAATCAAGATGATAAGCAAAATTTGGCATCATTAAC 924
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValIleuA 205
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
925 TTTCAAGGGGACGAATCTGTTAGTACGACATACATTACAAGAACAATGG 974
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
205 lalleIleAspAsnLysIleAsnProLeuSerLysAlaAspArgTyThr 221
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
975 AATTACAACCTGATCTTGTGGTGAATA...TGGGGAATAAATTTGAA 1021
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrlleuAs 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1022 GGTGCGCAATTCGAGAAGATTTGACGTCAATTCGCGATTATTATTAA 1071
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1072 TAATGGCTATGCCAAAGCAAAATTTACTAAACGGATGTTTCAGCTAAATG 1121

```

```

255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1122 ATGAAAAACAAAAAGTTAATGTAAACCATGTATGTAATGAAGTTTACAG 1171
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyThrGlnAl 288
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1172 TATGACCTTCGTAGTCACGATATTATAGTAAATCGAAGCAAAACTTCGAGAA 1221
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
288 aGluLeuGluAlaLeuLys...PheLysAlaGluGluGlyPheSerG 304
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1222 CGAGCTTGAACCTTTACTTTCAGCATTAATTAATGATACTTTCGCCGC 1271
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
304 lAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1272 GTAGTGATATTGCAGATGTAGAAAAATGCAATTAAGCAAAACTTCGAGAA 1321
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321 AspGlyTyTyThrTyAlaGlnIleArgProValThrArgIleAsnAspGl 337
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1322 CGCGTTTACGGTAGCGCAACGGTAAATTCAGTACCTGATTTTGATGATGC 1371
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 uSerArgThrValAspValGluTyTyThrIleAspProValHisProValT 354
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1372 AAATAAACATTTAGCGATAACCCCTTGTGTGATGCTGGACGACGTTTAA 1421
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1422 CTGTTCCGCAACTTCGCTTGAAGAAATACCGTTTCTGCTGATGACACT 1471
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnly 387
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1472 TTACGTGAGGAAATCGCCCAACAAGAGAACTTGGTATAATTCACAATT 1521
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 stIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPheLysHisV 404
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1522 AGTTGAGTTAGGAAAAATTCGCTTAGCTAGTACAGGTTTCTTCGAA... 1567
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
404 alThrValAspThrarg.....ProValProAsnSerProAspGlnVal 418
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1568 ..ACAGTCGAAACCGAATTTGATCTCAATGGTAGTAAATGATGAAGTG 1615
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1616 GATGCTGATATATAAGTCAAAAGAACGTAACACGGGTAGTATCAACTTGG 1665
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
435 aAlaGlyTySerGlnSerGlyValThrPheGlnPheAspValSerG 452
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1666 TATTGCTTACGCTACAGAGTGTATTAGTTATCAAGCAAGTGTATAAC 1715
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
452 lAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1716 AGATTAATTTCTTGGGAACAGGGCGGCAGTAGTATAGCTGTACGAAA 1765
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
469 SerGluThrArgGluValTySerLeuGlyMetThrAsnProTyThrPheTh 485
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1766 AATGATTATGCTACGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTAC 1815
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
485 rValAsnGlyValSerGlnSerLeuSerGlyTyTyThrArgLysThrLysT 502
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1816 TAAAGATGGTGA...AGTCTTGGTGAATGTTTCTTTTCAAAACT 1859
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
502 yrAspAsnLys.....AsnIleSerAsnTyValLeuAspSerTy 515
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1860 ACATTAACCTTAAAGTGATACATCTCTTAATTAAGCGTAGCAGCTTAC 1909
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
516 GlyGlySerLeuSerTyThrGlyTyThrProIleAspGluAsnGlnArgIleSe 532
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1910 GGAAGTAATGTTACTTTTAGGTTTCCCTGTAAATGAAATAACTCCTATTA 1959
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1960 TGTAGGATTAGTCACTATTAATAAATAGTAACCTTTGCTCTAGAAT 2009
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560

```

```

2010 ATAACCGTAATTATATATCAATCAATCAAA..... 2041
561 GlyLysIleGlnValAspAsnGlyIleProAspPheLysHisAspTy 577
2042 .....TTTAAAGGTAATGCGATTAAACAAATGACTTTGATTT 2079
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerLeuAspArgP 594
2080 TTCT.....TTTGGTTGGAACTATACACCCCTTAATAGAG 2114
594 roValPheProThrGlnGlyMetSerHisSerValasp.....LeuThr 608
2115 GCTATTTCCTCAACTAAAGGGTTAAAGCAAGTCTTGGTGACGAGTTACT 2164
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
2165 ATTCCAGGTTCTGTAAACAAATACTACAAACTAAAGTGCAGATGTACAGG 2214
625 eTyrArgProPhe.....IleLysLysSerValL 635
2215 TTCTACCATTTAGACAGAGATCACCTCTGGGTTGTATCTGCAAAAGCAT 2264
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn....AsnLeuProPheTyr 650
2265 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2311
651 GluAsnPheTyrAlaGlyCysTyrGlySerValArgGlyTyrAspGlnSe 667
2312 CAACTATTACAGCGGGTGGCATCGCTTACCTACGTTGGTTTGGTTATGG 2361
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
2362 TAGATTGGACCTAAGCA.....ATTATGCCGAATATGGTAATG 2402
684 lnThrThrLeuGly.....GluValValGlyGly 693
2403 GTAGTGGTACTGCTACTTTTAAAGAGATAAGTTCTGATGTGTTGGTGT 2452
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
2453 AATGCAATCGCTACAGCTAGCCAGAGTAATGTGCCAACCTCATTTGT 2502
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
2503 GAGCGATAGACCCAAATACGCTCCGACCTCTTATTGTGTATGCGG 2552
726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
2553 CAAGTGTTTGGAAATCTAAATGGAATCAGATAAAATGGATTAGAGAGC 2602
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAla 758
2603 GATGTATTAAAA..... 2614
758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
2615 .AGATTGCTGATTATGGCAATCAAGCGTATTCGCGCTCTACAGGTG 2663
775 aGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
2664 TCGGATTCCAATGGCAATCTCTATTGGCCCATTTGTATTCTCTTATGCC 2713
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
2714 AAACCAATTTAAAAATATGAAAAATGATGATGTCGAACAGTTCCAATTAG 2763
808 nIleGlySerValPhe 813
2764 TATTGGAGGTTCTTTC 2779

```

seq_name: /cgn2_6/ptodata/2/1na/6A_COMB.seq:US-09-135-166-55

```

seq_documentation_block:
: Sequence 55, Application US/09135166
: Patent No. 6083743
: GENERAL INFORMATION:
: APPLICANT: CHONG, Pele
: APPLICANT: THOMAS, Wayne
: APPLICANT: YANG, Yan Ping
: APPLICANT: LOOSMORE, Sheena
: APPLICANT: SIA, Dwo Yuan Charles
: APPLICANT: KLEIN, Michel
: TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
: NUMBER OF SEQUENCES: 55
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6TH Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/135,166
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/433,522
: FILING DATE: 12-SEP-1995
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: STEWART, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 55:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2987 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-09-135-166-55

alignment_scores:
Quality: 1052.50 Length: 822
Ratio: 1.956 Gaps: 20
Percent Similarity: 65.450 Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-09-135-166-55 ..
Align seg 1/1 to: US-09-135-166-55 from: 1 to: 2987

23 SerThrHisAlaGlnAlaAlaaspPheMetAlaAsnAspIleThrIleth 39
428 ACAACGACGTGTGTTTGGCCGACCTTTTGGCAAAAGATATTCGTGTGA 477
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
478 TGTGTGTCAAGGTGACTTGAACAACAAATCCGACCAAGTTTACCTGTTTC 527
56 rgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
528 GTGCCCGGTGAGGTGACTGACATGATGTGGTAATATTTGTCCTCTCT 577
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGlu 89
578 TTATTGCGTAAGTGGTTCGATTCGATGATGTGAAGCG...CATCAAGAAGG 624

```

89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
 122 CGATGCTGCTGTTGTTAGCGCTTGGCTAAATCGATCATTCAGATGTA 674
 105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
 675 AAATCAAGAGTAATCTGTTATTCCTCCACTGAAGCAGCTTAACAAACTTA 724
 122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG1 138
 725 GATCCTAACGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTA 774
 138 nMeIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrYrA 155
 775 TGAATTTGCCAAAAGTAAAGAGCAGCTATGCAAGTGTAGGTCGCTATA 824
 155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
 825 ACGCAACAGTTGAACCTATGCTCAATAGCTACCAATAATCGCGCTGAA 874
 172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
 875 ATTTAATTCAAATCAATGAAGATGATAAAGCAAAATTTGGCATCATTAAC 924
 188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
 925 TTTCAAGGGGAACCAATCTGTTAGCAGTACATTAACAGCAAAATGG 974
 205 laiIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
 975 AATTACAACCTGATCTTGGTGGAAATTA...TGGGGAATTAATTTGAA 1021
 222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
 1022 GGTGGCAATTCGAGAAAGATTTGCAGTCAATTCGTGATTATTATTAA 1071
 238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
 1072 TAATGGCTATGCCAAGACACAAATTAATAACGGATGTTTCAGCTAATG 1121
 255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
 1122 ATGAAAAAACAAGATTAATGTAACATGATGTAATGTAATGAAAGTTACAG 1171
 272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
 1172 TATGACCTTCGTAGTCAGCATTAATAGTAATCTGGGAGGTATCTCTGC 1221
 288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
 1222 CGAGCTTGAACCTTTACTTTTCAGCATTAATCAATTTAAATGATACTTTCCGCC 1271
 304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
 1272 GTAGTGATATTCGATGATAGAAATGCAATTTAAAGCAAAATTCGGAAA 1321
 321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG1 337
 1322 CCGGTTACGGTAGCGCAACGGTAAATTCAGTACCTGATTTGATGATGC 1371
 337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
 1372 AAATAAACAATTAGCATTAACCTGTTGTTGATGCTGGACGACGTTTAA 1421
 354 yValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
 1422 CTGTTGCCCACTTCGCTTTGAAGGAATACCGTTCTGCTGATGACACT 1471
 371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
 1472 TTACGTCAGAAATGCGCCCAACAAGAAAGCAACTTGGTATAATTCACAAT 1521

387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
 1522 AGTTGAGTTAGGAAAATTCGCTTAGATCGTACAGGTTTCTTCAA... 1567
 404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
 1568 ..ACAGTCGAAAACCGAATTCATCTATCAATGCTAGTAAATGATGAAGTG 1615
 419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
 1616 GATGTCGTATATAAAGCAAGACGTAACCGGTAGTATCAACTTGG 1665
 435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
 1666 TATTTGGTTACGGTACAGAGAGTGTATTAGTTATCAAGCAAGTGTAAAC 1715
 452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
 1716 AAGATAATTTCTTGGSAACAGCGCGCATAGTATAGCTGGTACGAAA 1765
 469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
 1766 AATGATATGTCAGCAGTGTCAATTTGGTGTATACCGAGCCCTATTATAC 1815
 485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
 1816 TAAAGATGTTGTA.....AGTCTTGGTGGAAATGTTTCTTTGAAAAC 1859
 502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
 1860 ACGATAACTCTAAAGATGATACATCTCTAATCTATAGCGTACGACTTAC 1909
 516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
 1910 GGAAGTAAATGTTACTTTAGTGTTCCTGTAATGAAATAACTCTCTATTA 1959
 532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
 1960 TGTAGGATTAGTGCATACCTATAATAAATTAGTAACCTTCTCTAGAA 2009
 544 LyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
 2010 ATAACCGTAATTTATATATCAATCAATGAAA..... 2041
 561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
 2042TTTAAAGGTAATGGCATTAATAACAAATGACTTTGATTT 2079
 577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgP 594
 2080 TTCT.....TTTGGTTGGAATATAACAGCCTTAATAGAG 2114
 594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
 2115 GCTATTTCCCAACTAAAGGGTTAAAGCAAGTCTTGGTGGAGAGTTACT 2164
 609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnI 625
 2165 ATTCAGGTTCTGTAAACAATACTACAACTAAGTGCAGATGTACAGGG 2214
 625 eTyrArgProPhe.....IleLysLysSerValL 635
 2215 TTCTACCATTAGACAGAGATCACCTCTGGGTTGTATCTGCAAAAGCAT 2264
 635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn.....AsnLeuProPheTyr 650
 2265 CTGCAGGATATGCAAA...GGTTTGGAAACAACGCTTACCGTTCTAT 2311
 651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
 2312 CAATCTATACAGCGGTGGCATCGGTCATTCATACGTTGTTTCTTATGG 2361
 667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684

```

:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2362 TAGATTGACCTAAGCA.....ATTATGCCGAATATGGTAATG 2402
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
684 lntThrThrLeuGly.....GluValValGlyGly 693
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2403 GTAGTGGTACTGCTACTTTTAAAGACATAAGTTCGTGATGTGATGGTGT 2452
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
694 AsnAlaLeuAlaThrPheGlySerGluLeuLeuLeuProLeuProPheLeu 710
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2453 AATGCAATCGCTACAGTACAGTACAGTAAATGTGCCCACTCCATTGCT 2502
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2503 GAGCGATAAGACCCAAATATACGCTCCGACCTCCTTATTGTTGATGCGG 2552
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
726 lYcInValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2553 CAAGTGTGTGGTAACTAAATGCAATCAGATAAATAAGATTAGAGACC 2602
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2603 GATGTATTAAAA..... 2614
758 asnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2615 AGATTGCTGATTATGCAAAATCAAGCGGTATTCGCGCTCTACAGGTG 2663
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 alGlyAlaThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAla 791
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2664 TCGGATTCATGGCAATCTCTATTGGCCATTGTTCTTCTTATGCC 2713
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2714 AAACCAATTTAAAAATATCAAAATGATGATGTCGAACAGTTCATTTAG 2763
808 nIleGlySerValPhe 813
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2764 TATGGAGGTTCTTTC 2779
```

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq.us-08-942-046-55

```

seq_documentation_block:
; Sequence 55, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 55:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2987 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-942-046-55

alignment_scores:
    Quality: 1052.50      Length: 822
    Ratio: 1.956          Gaps: 20
    Percent Similarity: 65.450      Percent Identity: 31.630

alignment_block:
US-09-701-711-2 x US-08-942-046-55 ..
Align seg 1/1 to: US-08-942-046-55 from: 1 to: 2987

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIlePh 39
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
428 ACAACAGCTGTGTGTCGCCACCTTTTGTGGCAAAAGATATTCGTGTGGA 477
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
478 TGGTGTCAAGGTGACTTAGAACACAAATCCGAGCAAGTTTACCTGTTC 527
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 rGlyGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 GTGCCGGTCAGCGTGTGACTGACATGATGTGGCTAATATTTGCCGCTCT 577
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGl 89
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
578 TTATTCGTAAAGTGGTCGATTCGATGATGTAAGCG...CATCAAGAGG 624
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
625 CGATGTGCTTGTGTAGCGTTGTGGCTAAATCGATCATTTTCAGATGTTA 674
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
675 AAATCAAGGTAACTCTGTTATTCCCACTGAAGCACTTAACAAAACTTA 724
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGl 138
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
725 GATGCTAACGGGTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATATAA 774
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 TGAATTTGCCAAAAGGTAAAGAGACACTATCAAGTGTAGTGCCTATA 824
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
825 ACGCAACAGTTGAACCTATTGTCAATACGCTACCAATAATCGCGCTGAA 874
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
875 ATTTTAATTCAAATCAATGAAGATGATAAGCAAAATTCGCATCAATTAAC 924
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
925 TTTCAAGGGGAACGAATCTGTTAGTAGCAGTACATTACAAGACAATGG 974
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
975 AATTACAACCTGATTCTTGGTGGAAATTA...TGGGAAATAAATTTGAA 1021
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

222 GluGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
 1022 GGTGCGCAATTCAGAAAGATTTTCAGCATCAATTCGTGATTTATTATAA 1071
 238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
 1072 TAATGGCTATGCCAACAACCAAAATTAATAACCGGATGTCACGTAATG 1121
 255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
 1122 ATGAAAAACAAGTTAATGTAACCATTCATGTAATGAAGGTTTACAG 1171
 272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
 1172 TATGACCTTCGTAGTCAGCAGCATTTATAGGTAATCTGGGAGGTATGCTGC 1221
 288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
 1222 CGAGCTTGAACCTTTACTTTCAGCATTTACATTTAAATGATACITTCGCGC 1271
 304 lAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
 1272 GTAGTGATATTGCAGATGTGAAAAATGCAATTAAGCAAAAACCTTGGAGAA 1321
 321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG1 337
 1322 CGCGGTACGGTACGCAACGGTAATTCAGTACTGATTTTGATGATGC 1371
 337 userArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
 1372 AAATAAAACATTTAGCGATACCCCTTGTGTGATGCTGGACGAGTTTAA 1421
 354 yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
 1422 CTGTTCGCAACTTCGTTTGAAGAAATACCGTTTCTGCTGATAGCACT 1471
 371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
 1472 TTACGTACAGAAATCGCCACAAGAGAACTTGGTATATATTCACAAT 1521
 387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHsv 404
 1522 AGTTGAGTTAGGAAAAATTCGTTTAGATCTAGATCTACAGGTTCTTCGAA... 1567
 404 alThrValAspThrArg...ProValProAsnSerProAspGlnVal 418
 1568 ..ACAGTCCGAACCGAATTCATCTATCAATGCTAGTAAATGATGAAGTG 1615
 419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
 1616 GATGTCGTATATAAGTCAAAAGAACGTAACACGGGTAGTATCAACTTGG 1665
 435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
 1666 TATTGGTTACGGTACAGAGGTGTTATAGTTATCAACGAAGTGTAAAC 1715
 452 lAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
 1716 AAGATAAATTCCTGGCAACAGGGCGGCAAGTAAGTATAGCTGGTACGAAA 1765
 469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
 1766 AATGATTATGTTAGGAGTGTCAATTTGGTTATACCGAGCCCTATTATTAC 1815
 485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
 1816 TAAAGATGGTGA.....AGTCTTGGTGGAAATGTTCTTTTGAAGAACT 1859
 502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
 1860 ACGATAACTCTAAAGTGTATACATCTCTAATCAATAAGCGTACGACTTAC 1909
 516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
 1910 GGAAGTAATGTTACTTAGTTCCTCGTAAATGAAATAAATCACTTATTA 1959
 532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
 1960 TGTAGGATTAGGTACATACCTATAATAAAATAGTAACCTTTGCTCTAGAAT 2009
 544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
 2010 ATAACCGCTAATTTATATATTCATCAATCAATGAAA..... 2041
 561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
 2042TTTAAAGCTAATGGCATTTAAACAAATGACTTTGATTT 2079
 577 rThrThrTyrAsnAlaIleLeuGlyTyrPasnTyrSerSerLeuAspArgP 594
 2080 TTCT.....TTTGGTTGGAATATACAGCCTTAATAAGAG 2114
 594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
 2115 GCTATTTCCCACTAAAGGGTTAAACCAAGTCTGTGGACGAGTTACT 2164
 609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnII 625
 2165 ATTCAGGTTCTGTATAACAAATACTACAACTAAGTGCAGATGTACAGGG 2214
 625 eTyrArgProphe.....IleLysLysSerValL 635
 2215 TTCTACCCCATTTAGACAGATACCTCTCGGTGTATCTGCAAAAGCAT 2264
 635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
 2265 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCCTCTAT 2311
 651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
 2312 CAACCTTTATACAGCGGTGCGTACGTTCACTAGCTGGTTTGTCTTATGG 2361
 667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
 2362 TAGTATGGACCTAACGCA.....ATTATGCCGAATATGTAATG 2402
 684 lNThrThrLeuGly.....GluValValGlyGly 693
 2403 GTAGTGGTACTGGTACTTTTAAAGATAAAGTCTCTGATGCTGATTGGTGT 2452
 694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuProPheLy 710
 2453 AATGCAATCGCTACAGTAGCGAGAGTTAATTTGGCCAACCTCCATTGT 2502
 710 sGlyAspTrpIle...AspGlnValArgProValIlePheIleGluGlyG 726
 2503 CAGCGATAAGACCAAAATACGGTCCGAACCTCTTATTTTGTGATGCGG 2552
 726 lyGlnValPheAspThrThr...GlyMetAspLysGlnThrIleAspLeu 741
 2553 CAAGTGTTTGGAATACTAAATGGAATCAGATAAAATGGATTAGAGAGC 2602
 742 ThrGlnPheLysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAl 758
 2603 GATGTATTAAAA..... 2614
 758 aAsnArgProLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyV 775
 2615 ..AGATTGCTGATTTATGCAAAATCAAGCGTATTCGCGCTCTACAGGTG 2663
 775 alGlyAlaThrTyrTyrThrProIleGlyProLeuSerIleSerTyrAla 791
 2664 TCGGATTCCAATGGCAATCTCTATTTGGGCCATTTGTTCTTATGCC 2713
 792 LysProLeuAsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGl 808

Thu Sep 19 10:04:26 2002

```

354 yrvAlaArgIleAsnPhetHrGlyAsnPhelYsThrGlnAspGluVal 370
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1423 CTGTTGGCAACTTCGCTTTGAAGAAATACCGTTCTGCTGATAGTACT 1472
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1473 TTACGTCAGGAAATCGCACAAAGAAAGAACTGGTATATAATTCACAATT 1522
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
387 stleGlnLeuSerArgAlaArgLeuMetArgHrGlyPhePheLysHisv 404
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1523 AGTTAGTTAGGAAATTCGTTAGATCGTACAGGTTTCTTCGAA... 1568
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1569 ..ACAGTTGAAACCGAATGTATCTCTATCAATGGTAGCAATGATGAAGTG 1616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
419 AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1617 GATGTCGTATATAAGTCAAGAACAACTAACACGGGTAGTATCAACTTTGG 1666
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
435 aLaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1667 TATTGTTAGCGGTACAGAGAGTGTATCAGTTTATCAACAAAGTATTAAAC 1716
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
452 lnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1717 AAGATATTTCTTGGGAACAGGGCGCGCAGTAGTAGTACGTGTCGAGAA 1766
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1767 AATGATTATGTCAGGAGTGTCAATTTGGTTATACCGAACCCCTATTATTAC 1816
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1817 TAAAGATGGTGTA.....AGTCTTGTGTGGAATATTTCTTTGAAAACT 1860
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
502 yAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1861 ACGATAACTCTAAAGATGATACCTCTACATATACGCGTACGACTAT 1910
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1911 GGAAGTAATGTTACTTTAGGTTTCCCTGTAATGAAATAACTCTCTATTA 1960
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1961 TGTAGGATTAGCCCATACCTATAATAAATTAGTAACCTTTGCTCTAGAAAT 2010
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
544 LyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2011 ATAACCGTAATTTATATATATATCAATCAATGAAA..... 2042
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2043 .....TTTAAAGGTAATGGCATTAATAACAAATAGCTTTGATTT 2080
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2081 TTCT.....TTTGGTTGGAATTAACACGCTTATATAGAG 2115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594 roValPheProThrGlnGlyMetSerHisSerValasp.....LeuThr 608
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2116 GCTATTTCCCAACTAAGGGTTAAAGCAAGTCTTGGTGGACGAGGTACT 2165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
609 valGlyPheGlyAspLysThrHisGlnLysValVal.....TyrGlnG 623
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2166 ATTCCAGGTTCTGATAACAAATATACAACTAAGTGCAGATGTACAGG 2215
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
623 y.....AsnIleTyrArgProPheIleLysLysSerValL 635
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2216 TTTCACCATTAGACAGAGATCACCGCTGGTGTGTATCTGCAAAAGCAT 2265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-09-135-166-9

```

seq documentation block:
; Sequence 9, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```



```

alignment_scores:
  Quality: 1047.50      Length: 818
  Ratio: 1.947          Gaps: 21
  Percent Similarity: 65.770  Percent Identity: 31.663

alignment_block:
  US-09-701-711-2 x US-09-135-166-9  ..

  Align seg 1/1 to: US-09-135-166-9  from: 1 to: 2989

23  SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
429 ACAACGACTGTTGTTGCCGACACCTTTTGTGCCAAAAAGATATCGTGGA 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

39  rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
479 TGGTGTTCAGGTCACCTTAGAACACAAATCCGAGCAAGTTTACCTGTTC 528
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

56  rgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
529 GTGCTGTCAGCGCTGTACGACAAATGATGTGCTAATATTGTCCGCTCT 578
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

73  LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
579 TTATTCGTAAGTGTGCTGATTCGATGATGTGAAGCG...CATCAAGAAGG 625
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

89  yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
626 CGATGTCGCTGTTGTTAGCGCTGTGGCTAATTCGATCATTTTCAGATGTTA 675
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

105  snPheGluGlyAsnArgLeuIleProLysGlnGlyLeuGlnGluGlyLeu 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
676 AAATCAAGGTAACTCTGTTATTTCCCTAGACACTTAACAAAAACCTTA 725
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

122  LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
726 CATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAAAATTTAA 775
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

138  nMeIleGluThrCluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
776 TGAATTTGCCAAAGTGTAAAGAGCACTATGCAAGTGTAGGTCGCGTATA 925
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

155  snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1767 AATGATTATGGTAGCAGTGCAATTTGGGTATATACCGAACCTATTATAC 1816
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1817 TAAAGATGGTGTA.....AGTCTGGTGGAAATATTTCTTTGAAACT 1860
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
502 YrAspAsnLys.....AsnLysSerAsnTyrValLeuAspSerTyr 515
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1861 ACGATAACTCTAAAGTAGATACATCTCTAACTATAGCGGTACGACTTAT 1910
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
516 GlyGlySerLeuSerTyrGlyTyrProLysPheGlnAsnGlnArgLys 532
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1911 GGAAGTAAATGTTACTTTAGTGGTTCCTCTAAATGAAATAACTCTCTATTA 1960
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1961 TGTAGGATTAGCCCATACCTATATAAATTAGTAACCTTTGCTCTAGAAT 2010
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
544 lyGlyArgPheMetGlyLysSerAsnValLysGlnLeuMetAlaAspGly 560
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2011 ATAACCGTAATTATATATATCAATCAATCAAA..... 2042
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
561 GlyLysLleGlnValAspAsnGlyLysProAspPheLysHisAspTy 577
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2043 .....TTTAAAGGTAATGGCAATTAACAAATGACTTTGATTT 2080
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
577 rThrThrTyrAsnAlaLleLeuGlyTyrAsnTyrSerSerLeuAspArgP 594
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2081 TTCT.....TTTGGTTGGAACCTATAACAGCCTTAATAGAG 2115
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2116 GCTATTTCCTCAACTAAAGGGTAAAGCAAGCTTGGTGGAGAGTTACT 2165
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
609 ValGlyPheGlyAspLysThrHisGlnLysValVal.....TyrGlnG 623
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2166 ATCCAGGTTCTGATAACAACTACTACAACTAAGTGAGATGTACAGG 2215
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
623 Y.....AsnLysTyrArgProPheLysLysSerValL 635
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2216 TTCTACCATTTAGACAGAGATCACCGCTGGTGTGTATCTGCAAAAGCAT 2265
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn....AsnLeuProPheTyr 650
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2266 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2312
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
651 GluAsnPheTyrAlaGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2313 CAACCTATACAGCGGTGGCATGTGTTCTATACCGGTTTTCCTATGG 2362
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 584
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2363 TAGATTGGGCTCAATGCA.....ATTATGCCGAACATGTTAATG 2403
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
684 lnThr.....ThrLeuGlyGluValValGlyGlyAsnAlaLeuAla 697
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2404 GTACTTTTAAATAGATAAGTTCTGATGTATGTTGGTAAATGCAATCACA 2453
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
698 ThrPheGlySerGluLeuLeuProLeuProPheLysGlyAspTrpIl 714
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2454 ACTGCAGGTGCAGAACTTATTGTACCAACTCCATTTGTGAGTGATAAAG 2503
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
714 e....AspGlnValArgProValLlePheLleGluGlyGlyGlnValPhe 730
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2504 CCAAAATACAGTCCCAACCTCCCTATTGTGTGATGCGCAAGTGTGGGA 2553
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
730 spThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPheLys 745
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2554 ATACTAAATGGAAATCAGATAAATAATGGATTAGAGAGCAAGGCTTGAAA 2603
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
746 AspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLe 762
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2604 GAC.....TTACCTGA 2614
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
762 uLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrT 779
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2615 TTATGGCAAAATCAAGCGTATTCGCGCTCTACAGGTGTCGGATTCCAAT 2664
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
779 rPtyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsn 795
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2665 GCGAATCTCTATTGGACCATTTGTTATTTCTTATGCTAAACCAATAAA 2714
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
796 LysLysGlnAsnAspGlnThrAspThrValGlnPheGlnLleGlySerVa 812
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2715 AAATATGAAATGATGATGTGCAACAGTTCCAATTTAGTATTGGGGCTC 2764
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
812 lPhe 813
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
2765 TTTC 2768
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:us-08-942-046-9
seq_documentation_block:
; Sequence 9, Application US/08942046
; Patent No. 6264954
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Iuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/942,046
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-732 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 390..2768
; US-08-942-046-9
```

alignment_scores:
 Quality: 1047.50 Length: 818
 Ratio: 1.947 Gaps: 21
 Percent Similarity: 65.770 Percent Identity: 31.663

alignment_block:
 US-09-701-711-2 x US-08-942-046-9 ..

Align seg 1/1 to: US-08-942-046-9 from: 1 to: 2989

```

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
429 ACAACGACTGTGTTGGCCGACCTTTGGCCCAAGATATTCGTGTGA 478
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
479 TGGTGTTCAGGTGACTTAAAGACAAATCCGAGCAAGTTTACCTGTTTC 528
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
56 rGlyLeuGlnValValSerGluAsnGlnLeuAlaAspValValLysAla 72
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
529 GTGCTGGTCAGCGTGTGACTGACATGATGTGCTAAATATTCGCGCTCT 578
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG 89
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
579 TTATTCGTAAAGTGTGATTCGATGTGAAAGCG...CATCAAGAAGG 625
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
626 CGATGTGCTGTGTTGTTAGCGTTAAATTCGATCATTTTCAGATGTTA 675
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
676 AAATCAAGGTAACTCTGTATTCCTCCACTGAAGCACTTAACAAACTTA 725
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
122 LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValG 138
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
726 GATGCTAACGGGTTTAAAGTTGGCGATGTTTAAATTCGAGAAATTA 775
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
138 nMetIleGluThrGluThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
776 TGAATTTGCCAAAGTGTAAAGACACTATCCAAAGTGTAGTGTGCTATA 825
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
826 ACGCAACCGTTGAACCTATTGTCAATACGTCGCCAAATAATCGTGTGAA 875
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
172 LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
876 ATTTTAATTCAAATCAATGAAGATGATAAAGCAAAATTTGGCATCATTA 925
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
926 TTTCAAGGGGAACGAATCTGTTAGTAGCAGTACATTACAGAACAAATGG 975
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
976 AATTACAACCTGATCTTGGTGGAAATTA...TGGGAAATAAATTTGAA 1022
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1023 GGTGGCGCAATTCGAAAGATCTGAGGCAATTCGTGATTATTATTATAA 1072
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1073 TAATGGCTATGCCAAGGACAAATCACTAAACGGATGTTTCAGCTAAATG 1122
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGlnGln 271
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1123 ATGAAAAACAAAGCTTAATGTAACTGATGTGTAATGAAGTTTACAG 1172
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

1173 TATGACCTTCGTTAGTCCACGCATTATAGTAATCTCGGAGGTATGTCTCC 1222
288 aGluLeuGluAlaLeuLeuLys...PheLysAlaGluGluGlyPheSerG 304
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1223 CGAGCTTGAACCTTTACTTTTCAGCATTTACATTTAAATGATACTTTCCGCC 1272
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
304 InAlaMetLeuGluGlnThrAsnAsnIleSerThrLysPheGlyAsp 320
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1273 GTAGTGATATTTCAGATGTAGAAATGCAATTAAGCAAAACTTTGGGAA 1322
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG 337
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1323 CGAGGTACGGTAAACACACAGATAAATCTGTACCTGATTTTGACGATGC 1372
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1373 AAATAAAACATTAGCGATAAATCTTGTGTTGATGCTGCGACGACGTTTAA 1422
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
354 yfValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1423 CTGTTCCGCAACTTCGCTTTGAAGGAATACCGTTTCTGCTGATAGTACT 1472
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1473 TTACGTCAGGAAATCGACACAAGAAGAACTTGGTATAAATTCACAATT 1522
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1523 AGTTGAGTTAGGAAATTCGCTTAGATCGTACAGGTTTCTTCGAA... 1568
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
404 alThrValAspThrArg...ProValProAsnSerProAspGlnVal 418
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1569 ..ACAGTTGAAACCGAATTTGATCTATCAATGGTAGCAATGATGAAGTG 1616
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1617 GATGTCGTATATAAAGTCAAGAAGCTAACCGGTTAGTATCAACTTTGG 1666
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1667 TATTTGTTACGGTACAGAGTGTCAATTTGGGTTATACCGAACCCCTATTAC 1716
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1717 AAGATAATTTCTTGGACACAGGCGGCAGTAGTAAATAGTATAGCTGTA 1766
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1767 AATGATTATGGTACGAGTGTCAATTTGGGTTATACCGAACCCCTATTAC 1816
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
485 fValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1817 TAAAGATGGTGTA...AGTCTTGGTGGAAATATTTTCTTTGAAAACT 1860
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
502 yfAspAsnLys...AsnIleSerAsnTyrValLeuAspSerTyr 515
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1861 ACGATAACTCTAAAAGTGTATACATCTCTAATAAGCGTACGACTTAT 1910
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1911 GGAAGTAATGTACTTTAGGTTTCCCTGTAAATGAAATAAATAACTCTAT 1960
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
532 rPheGlyLeu...AsnAlaAspAsnThrLysLeuHisG 544
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
1961 TGTAGATTAGCCCATACCTATATAATAATTAATAGTAACTTTGCTCTAGA 2010
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
2011 ATAACCGTAATTTATATATATCAATCAATGAA... 2042
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
2043 .....TTTAAAGGTAATGGCATTAACAAACAAATGACTTTGATT 2080
   |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```
577 rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2081 TTCT.....TTTGGTTGGAActATAACAGCCTTAATAGAG 2115
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
594 roValPheProThrGlnGlyMetSerHisSerValasp.....LeuThr 608
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2116 GCTATTTCACCACTAAAGGGGTTAAAGCAAGTCTTGGTGACGAGTTACT 2165
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
609 ValGlyPheGlyAspLysThrHisGlnLysVal.....TyrGlnG1 623
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2166 ATTCAGGTTCTGTATACAAATACTACAAATACTAGTCAGATGTACAGGG 2215
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
623 Y.....AsnIleTyrArgProPheIleLysLysSerValL 635
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2216 TTTCTACCATAGACAGATACCCCTGGTGTATCTGCAAAAGCAT 2265
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2266 CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTTCTAT 2312
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
651 GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2313 CAAACTTATACAGCGGTGGCATTTGGTTCATTACGCGGTTTGGCTTATGG 2362
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2363 TAGTATTGGGCTTAATGCA.....ATTATGCCGACATGGTAATG 2403
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
684 lnThr.....ThrLeuGlyGluValValGlyGlyAsnAlaLeuAla 697
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2404 GTACTTTTAATAAGATAAGTTCTGATGTGTTGGTGAATGCAATCACA 2453
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
698 ThrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTrpI1 714
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2454 ACTCGAGTGCAGAACCTATTGTACCAACTCCATTTGTGAGTGATAAAG 2503
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
714 e...AspGlnValArgProValIlePheIleGluGlyGlyGlnValPheA 730
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2504 CCAAAATACAGTCCGACCTCCCTATTGTGTGATCGCGCAAGTGTGGA 2553
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
730 spThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPheLys 745
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2554 ATACTAAATGGAATACAGATAAAATGGATTAGAGAGCAAGGTCTTGAAA 2603
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
746 AspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgProLe 762
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2604 GAC.....TTACCTGA 2614
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
762 uLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaThrT 779
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2615 TTATGGCAAAATCAAGCGGTATTCGCGCTCTACAGGTGTCGGATTCCAAT 2664
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
779 rPtyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeuAsn 795
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2665 GGCAATCTCTATTGGACCATTTGTTTCTTATGCTAAACCAATTTAA 2714
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
796 LysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySerVa 812
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2715 AAATATGAAATGATGATGTGCAACAGTTCCAATTTAGTATTGGGGCTC 2764
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
812 lPhe 813
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
2765 TTTC 2768
```

seq_name: /cgn2_6/prodata/2/ina/6A_COMB.seq:US-08-433-522A-7

seq_documentation_block:

; Sequence 7, Application US/08433522A
; Patent No. 6013514
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne

```
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,522A
; FILING DATE: 12-SEP-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 386..2761
; US-08-433-522A-7
```

alignment_scores:

Quality: 1042.00 Length: 819
Ratio: 1.940 Gaps: 20
Percent Similarity: 65.568 Percent Identity: 31.380

alignment_block:

US-09-701-711-2 x US-08-433-522A-7 ..

Align seg 1/1 to: US-08-433-522A-7 from: 1 to: 2974

```
23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
425 ACAACGACTGTGTTCGCCACCTTTTGTGGCAAAAGATATTCTGTGGA 474
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
475 TGTGTTCGAAGGTGACTTAGAACAAACAATCCGAGCAAGTTTACCTGTC 524
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
56 rGlyGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
525 GTCCCGTCAGCGGTGCTGACTGACAATGATGTGCTAATATTGTCGCTCT 574
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG1 89
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
575 TTATTCGTAAGTGTGTCGATTCGATGATGTGAAGCG...CATCAAGAGG 621
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
89 YArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluLea 105
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
622 CGATGTGCTGTTGTTGTAGCGTTGTGGCTAAATCGATCATTTTCAGATGTA 671
    :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
```

```

..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
672  AATCAAAGGTAACCTCTATTATCCACCTGAAGCAGCTAAACAAACTTA 721

122  LysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrValGI 138
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
722  GATGCTAACGGGTTTAAAGTTGGCGATATTTAAATTCGAGAAAAATTTAA 771

138  nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrYrA 155
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
772  TGAATTTGCCCAAGTGTAAGAGCAGCAGCTATGCAAGTGTAGGTCGCTATA 821

155  snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
822  ACGCAACCGTTGAACCTATTGTCAATACGCTACCAATAATCGCGCTGAA 871

172  LeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIleAs 188
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
872  ATTTTAATCAATCAATGAAGATGATAAAGCCCAATTTGGCATCATTAAC 921

188  nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
922  TTCAAGGGGACGAATCTGTAGTACAGTACATTTACAAGAACAAATGG 971

205  laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
972  AATTACAACCTGATCTCTGTGGTGAATTA...TGGGGAATAAATTTGAA 1018

222  GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1019  GTGCGCAATTCGAGAAAGATTTGCGGCAATTTGCGTATTATTATTTAA 1068

238  nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1069  TAATGGGTATGCCAAGACACAATCACTAAAGCGGATGTTTCAGCTAATG 1118

255  luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGln 271
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1119  ATGAAAAACAAAAAGTTAATGTAACCATTTGATGTAATGAAGTTTACAG 1168

272  TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1169  TATGACCTTCGTAGTCACCATATAGGTAATCTGGGAGGTATGCTGTC 1218

288  aGluLeuGluAlaLeuLys...PheLysAlaGluGluGlyPheSerG 304
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1219  CGAGCTGAACCTTTACTTTTCAGCATTTACATTTAAATGATCTTCGCC 1268

304  lnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1269  GTAGTGATATTGCAGATGTAGAAAAATGCAATTAAGCAAAACTTGGGGAA 1318

321  AspGlyTyrTyrThrAlaGlnIleArgProValThrArgIleAsnAspGI 337
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1319  CGAGGTTCAGGTAAACACACAGTAAATTCGTATCTGATTTGACGATGC 1368

337  uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1369  AATAAACAATTAAGCATACCTTTGTTGTTGATGCTGGAGCGGTTTAA 1418

354  yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1419  CTGTTCAACCAACTTCGCTTTGAAGGAATACCGTTTCTGCTGATAGTACT 1468

371  LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1469  TTACGTGAGGAAATGCGCCCAACAGAGGAACCTTGGTATATATTCACAATT 1518

387  sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1519  AGTTGAGTTAGGAAATTCGTTAGATCGTACAGGTTTCTTCGAA... 1564

404  alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||

```

```

1565  ..ACAGTTGAAACCGAATTGATCCTATCAATGCTAGCAATGATGAAGTG 1612

419  AspValAsnPheValValGluGluGlnProSerGlySerSerThrIleAl 435
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1613  GATGCTGATATAAAGTCAAGAACGTAACACGGGTAGTATCAACTTTGG 1662

435  aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1663  TATTTGGTTACGTACAGAGAGTGTATTAGTTATCAAGCAAGTGTCAAC 1712

452  lnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1713  AAGATAATTTCTTGGGAACGCGGCGAGTAAGTATAGCTGGTACGAAA 1762

469  SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1763  AATGATTATGATGAGTGTCAATTTGGTTATACCGAGCCCTATTTTAC 1812

485  rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1813  TAAAGATGGTGA.....AGTCTGGTGGAAATGTTTCTTTGAAAAACT 1856

502  yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1857  ACGATAACTCTAAAGGTGATACATCTCTAACTATTAAGCGTACGACTTAT 1906

516  GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1907  GGAAGTATGTTACTTTAGGTTTCCCTGTAATAAGAAATAACTCCTATTA 1956

532  rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
1957  TGTAGGATTAGGCCATACCTATATAATAATAGTAACTTTCCTCTAGAA 2006

544  lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2007  ATACCGTAATTTATATATATCAATCAATGAAA..... 2038

561  GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2039  .....TTTAAAGGTAATGGCATTAACAAATAGACTTTTGATTT 2076

577  rThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSerLeuAspArgp 594
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2077  TTCT.....TTTGGTTGAACTATAACAGCCTTAATAGAG 2111

594  roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2112  GCTATTTCCCAACTAAAGGGGTTAAAGCAAGCTTTGGTGGACGAGTTACA 2161

609  ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnII 625
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2162  ATTCAGGTTCTGTATACAAATACTACAACTAAGTGCAGATGTACAGGG 2211

625  eTyrArgProPhe.....IleLysLysSerValL 635
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2212  TTCTACCCATTAGACAGATCACTCTGGTTGTATCTCCTCAAAAGCAT 2261

635  euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2262  CTGCAGGATATGCAAT...GGTTTGGAAACAAGCGTTTACCGTCTCTAT 2308

651  GluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2309  CAAACTTATACAGCGGTGCTGTTGTTTCATTACGCGGTTTGTCTTATGG 2358

667  rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2359  TAGCATTTGGCCCTAACCA.....ATTATCATCAGGTCAAA 2393

684  lnThrThrLeu.....GlyGluValValGlyGlyAsnAlaLeu 696
: ..... ||||| ..... ||||| ..... ||||| ..... ||||| ..... |||||
2394  ATAATAAATTTAATAAGATAAGTCTGATGATTGTTGGTGAATGCAATC 2443

```

```

697 AlaThrPheGlySerGluLeuLeuLeuProLeuProPheLysGlyAspThr 713
||||| : : : : : ||||| : : : : : ||||| : : : : :
2444 GCTACAGTAGCGCAGAGTAATTTGGCAACTCCATTGTGAGTGATAA 2493
713 p1e...AspGlnValArgProValIlePheIleGluGlyGlnValP 729
: : : : : ||||| : : : : : ||||| : : : : : |||||
2494 GAGTCAAAATACAGTCCGAACTCCCTATTCTTGTATGCGCAAGTGT 2543
729 heAspThrThr...GlyMetAspLysGlnIleAspLeuThrGlnPhe 744
: : : : : ||||| : : : : : ||||| : : : : : |||||
2544 GGAATACTAAATGGAATCAGATAAAATGGATTAGAGAGCAATGCTTG 2593
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
||||| : : : : : ||||| : : : : : ||||| : : : : :
2594 AAAGAC.....TTACC 2604
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 778
| : : : : : ||||| : : : : : ||||| : : : : : |||||
2605 CGATTATGCAATCAAGCGTACTCGCGCTCTACAGGTGTCGGATTCC 2654
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
||| : : : : : ||||| : : : : : ||||| : : : : : |||||
2655 AATGGCAATCTCTAGTAGCAGCAGTGGTATTTCTATGCTAAACCAAT 2704
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
: : : : : ||||| : : : : : ||||| : : : : : |||||
2705 AAAAAATATGAAATGATGATCGAAGAGTCCCAATTTAGTATTGGGGG 2754
811 rValpHe 813
2755 TTCTTTC 2761

```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-135-166-7

```

seq_documentation_block:
; Sequence 7, Application US/09135166
; Patent No. 6083743
; GENERAL INFORMATION:
; APPLICANT: CHONG, Pele
; APPLICANT: THOMAS, Wayne
; APPLICANT: YANG, Yan Ping
; APPLICANT: LOOSMORE, Sheena
; APPLICANT: SIA, Dwo Yuan Charles
; APPLICANT: KLEIN, Michel
; TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6TH Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/135,166
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/433,522
; FILING DATE: 12-SEP-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: STEWART, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-829 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155

```

```

; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2974 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 386..2761
; US-09-135-166-7

```

```

alignment_scores:
  Quality: 1042.00      Length: 819
  Ratio: 1.940          Gaps: 20
Percent Similarity: 65.568 Percent Identity: 31.380
alignment_block:
US-09-701-711-2 x US-09-135-166-7 ..

```

Align seq 1/1 to: US-09-135-166-7 from: 1 to: 2974

```

23 SerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrIleTh 39
: : : : : ||||| : : : : : ||||| : : : : : |||||
425 ACAAGGACTGTGTTTGGCGCACCTTTTGGCAAAAGATATTTCGTGTGA 474
39 rGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuProPheA 56
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
475 TGGTGTTCAGGTGACTTAGACACACAATCCGAGCAAGTTTACCTGTTTC 524
56 rGluGlyGlnValValSerGluAsnGlnLeuAlaAspGlyValLysAla 72
|| : : : : : ||||| : : : : : ||||| : : : : : |||||
525 GTGCCGGTCAGCGTGTGACTGACAAATGATGTGGTAATATTGTCCGCTCT 574
73 LeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluG 89
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
575 TTATTCGTAAGTGGTGCATTCGATGATGAAAGCG...CATCAAGAAGG 621
89 yArgIle...IleTyrGlnValThrGluArgProLeuIleAlaGluIleA 105
| : : : : : ||||| : : : : : ||||| : : : : : |||||
622 CGATGTGCTTGTGTAGCGTTGTGGCTAAATCGATCATTTCCAGATGTTA 671
105 snPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyLeu 121
: : : : : ||||| : : : : : ||||| : : : : : |||||
672 AAATCAAGGTAACCTCTATTATTCCACCTGAAGCAGCTATAAACAACCTTA 721
122 LysAsnAlaGlyLeuAlaValGlyGlnProLysGlnAlaThrValG 138
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
722 GATGCTAACGGGTTTAAAGTTGGCGATATTTAATTCGAGAAAATATAA 771
138 nMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrA 155
: : : : : ||||| : : : : : ||||| : : : : : |||||
772 TGAATTTGCCCAAAGTGTAAGAGAGACATATGCAAGTGTAGTTCGCTATA 821
155 snThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValLys 171
||||| : : : : : ||||| : : : : : ||||| : : : : : |||||
822 ACGCAACCGTTGAACCTATTGTCAATACGCTACCAATAATCGCGTGAA 871
172 LeuAspMetThrPheAlaGlyLysProAlaArgValValAspIleAs 188
: : : : : ||||| : : : : : ||||| : : : : : |||||
872 ATTTTAATTCAAATCAATGAAGATGATAAGCCAAATGGCATCATTAAC 921
188 nIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLeuA 205
: : : : : ||||| : : : : : ||||| : : : : : |||||
922 TTTCAGGGGGAACGAATCTGTTAGTAGCAGTACATTACAAGAACAATGG 971
205 laIleLysAspAsnLysIleAsnProLeuSerLysAlaAspArgTyrThr 221
: : : : : ||||| : : : : : ||||| : : : : : |||||
972 AATTCAACCTGATTCTTGGTGGAAATTA...TGGGAAATAAATTGAA 1018
222 GlnGluLysLeuValThrSerLeuGluAsnLeuArgAlaLysTyrLeuAs 238

```

```
1019 GGTCCGCAATTCGAGAAGATTTCAGGCAATTCGTGATTATTATTTAAA 1068
      ::::: ::::: ::::: ::::: ::::: :::::
238 nAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnIleAsnG 255
      |||::: ::::: ::::: ||| ::::: :::::
1069 TAATGGCTATGCCAAGACACAAATCACTAAAGCGGATGTTTCAGCTAAATG 1118
      ::::: ::::: ::::: ::::: ::::: :::::
255 luAspLysAsnArgIlePheValGluIleSerLeuHisGluGlyGluGln 271
      ::::: ::::: ::::: ||| ::::: ::::: |||
1119 ATGAAAAACAAAAGTTAATGTAACTTATGTTAAATGAAGGTTTACAG 1168
      ::::: ::::: ::::: ||| ::::: ::::: |||
272 TyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeuThrTyrThrGlnAl 288
      ||| ::::: ::::: ::::: ||| ::::: :::::
1169 TATGACCTTCGTAGTCAGCAGCATATAGGTAATCTGGGAGGTATGCTGC 1218
      ::::: ::::: ::::: ::::: ::::: :::::
288 aGluLeuGluAlaLeuLys...PheLysAlaGluGluGlyPheSerG 304
      |||::: |||::: |||::: |||::: |||::: |||:::
1219 CGAGCTTGAACCTTTACTTTCAGCATTTACATTTAAATGATACTTTCGCC 1268
      ::::: ::::: ::::: ::::: ::::: :::::
304 InAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAsp 320
      ::::: ::::: ::::: ||| ::::: ::::: |||
1269 GTAGTGATATGCAGATGTAGAAAATCAATTAAGCAAACTTGGGAA 1318
      ::::: ::::: ::::: ||| ::::: ::::: |||
321 AspGlyTyrTyrTyrAlaGlnIleArgProValThrArgIleAsnAspG 337
      |||::: |||::: |||::: |||::: |||::: |||:::
1319 CGAGGTTACGGTAACAACAAGTAAATCTGTACCTGATTTGACGATGC 1368
      ::::: ::::: ::::: ||| ::::: ::::: |||
337 uSerArgThrValAspValGluTyrTyrIleAspProValHisProValT 354
      ::::: ::::: ::::: ||| ::::: ::::: |||
1369 AAATAAAACATATACGGTAACTTGTGTTGATGCTGGACGAGTTTAA 1418
      ::::: ::::: ::::: ||| ::::: ::::: |||
354 yrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAspGluVal 370
      |||::: |||::: |||::: |||::: |||::: |||:::
1419 CTGTTCCCAACTTCGTTGAAGGAATACCGTTTCTGCTGATGATGACT 1468
      ::::: ::::: ::::: ||| ::::: ::::: |||
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLy 387
      |||::: |||::: |||::: |||::: |||::: |||:::
1469 TTACGTCAGCAATTCGCCAACAAGAACTTGGTATATTAATCACAAAT 1518
      ::::: ::::: ::::: ||| ::::: ::::: |||
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
      ::::: ::::: ::::: ||| ::::: ::::: |||
1519 AGTTGAGTTAGGAAAATTCGCTAGATCGTACAGGTTTCTCGAA... 1564
      ::::: ::::: ::::: ||| ::::: ::::: |||
404 alThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
      |||::: |||::: |||::: |||::: |||::: |||:::
1565 ..ACAGTTGAAACCGAATTCCTCTCAATGATGACCAATGATGAAGTG 1612
      ::::: ::::: ::::: ||| ::::: ::::: |||
419 AspValAsnPheValValGluGlnProSerGlySerSerThrIleAl 435
      |||::: |||::: |||::: |||::: |||::: |||:::
1613 GATGTCGTATATAAGTCAAAAGCAAGCAACACGGTAGTATCAACTTTGG 1662
      ::::: ::::: ::::: ||| ::::: ::::: |||
435 aAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspValSerG 452
      |||::: |||::: |||::: |||::: |||::: |||:::
1663 TATTTGTTTACGACAGAGTGGTATGTTATATACAGCAAGTGTCAAC 1712
      ::::: ::::: ::::: ||| ::::: ::::: |||
452 InAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
      |||::: |||::: |||::: |||::: |||::: |||:::
1713 AGATAATTTCTTGGGNAACAGGGCGGCAGTAGTATAGTGGTACGAAA 1762
      ::::: ::::: ::::: ||| ::::: ::::: |||
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
      ::::: ::::: ::::: ||| ::::: ::::: |||
1763 AATGATTTATGTCAGAGTGCAATTTGGGTTATACCGAGCCCTATTTTAC 1812
      ::::: ::::: ::::: ||| ::::: ::::: |||
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysT 502
      |||::: |||::: |||::: |||::: |||::: |||:::
1813 TAAAGATGGTGA.....AGTCTGGTGGAAATGTTTCTTTGAAAACT 1856
      ::::: ::::: ::::: ||| ::::: ::::: |||
502 yrAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
      |||::: |||::: |||::: |||::: |||::: |||:::
1857 ACGATACTCTAAAGTATACATCTCTCACTAATAAGCGTACGACTAT 1906
      ::::: ::::: ::::: ||| ::::: ::::: |||
516 GlyClySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
      |||::: |||::: |||::: |||::: |||::: |||:::
```



```
1369 AAATAAACACATTAGCAGTAAACCTTTGTTGTTGATGCTGGACGAGCTTTAA 1418
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
354 yValargArgIleAsnPhThrGlyAsnPhelysthrGlnAspGluVal 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1419 CTGTTCCCAACACTTCGCTTTGAAGAAATACCCGTTTCTGCTAGTAGTAT 1468
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
371 LeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLys 387
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1469 TTACGTCAGAAATGCGCCCAACGAAGGAACCTTGGTATAATATCAACAAT 1518
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
387 sIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePheLysHisV 404
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
1519 AGTTGAGTTAGGAAAAATTCGCTTAGCTAGCTACAGGTTTCTTCGAA... 1564
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
404 aIThrValAspThrArg.....ProValProAsnSerProAspGlnVal 418
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1565 ..ACAGTTGAAAACCGAAATGTGATCTATCAATGGTAGCAATGATGAAGTG 1612
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
419 AspValAsnPhValValGluGluGlnProSerGlySerSerThrIleAl 435
      ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1613 GATGTCGTATATAAGATCAAGAACGTACACGGGTAGTATCAACTTTGG 1662
      ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
435 aAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerG 452
      : ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: :: ::
1663 TATTGTTACGGTACAGAGAGTGCTATTAGTTATCAAGCAAGTGTCAAAC 1712
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
452 lnAsnAsnPhMetGlyThrGlyLysHisValAsnAlaSerPheSerArg 468
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1713 AGATATAATTTCTGGGAACAGGGCGGCAGTAGTATAGTGTGTCAGGAA 1762
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
469 SerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheTh 485
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
1763 AATGATATGTTAGCAGAGTGTCAATTTGGGTTATACCGAGCCCTATTTTAC 1812
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
485 rValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLysL 502
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1813 TAAAGATGGTGTA....AGTCTGGTGGAATGTTTCTTTGAAAACT 1856
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
502 yRAspAsnLys.....AsnIleSerAsnTyrValLeuAspSerTyr 515
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1857 ACATAACTCTAAAGTGATACATCTCTAACTATAAGCGTAGCAGCTTAT 1906
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
516 GlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSe 532
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1907 GGAAGTAATGTTACTTAGTTCCTCTGTAAGTGAATAAATCACTCTATTA 1956
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
532 rPheGlyLeu.....AsnAlaAspAsnThrLysLeuHisG 544
      ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: ::
1957 TGTAGGATTAGGCATACCTATAATAATAAATTAGTAATTTGCTCTAGAAT 2006
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
544 lyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspGly 560
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
2007 ATACCGTAATTATATATATTAATCAATCAATGAA..... 2038
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
561 GlyLysIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTy 577
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
2039 .....TTTAAAGGTAATGCACTATAAAACAACAAATGACTTTGATT 2076
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
577 rThrThrTyrAsnAlaIleLeuGlyTrpAsnTyrSerSerLeuAspArgP 594
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
2077 TTCT.....TTTGGTTGGAACATATAACAGCCTTAATAGAG 2111
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
594 roValPheProThrGlnGlyMetSerHisSerValAsp.....LeuThr 608
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2112 GCTATTTCCCACTAAAGGGGTTAAAGCAAGTCTTGGTGACGAGTTACA 2161
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
609 ValGlyPheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIl 625
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
2162 ATTCCAGGTTCTGATAACAATACTACAACTAAGTGCAGATGTCACAGG 2211
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
625 eTyrArgProPhe.....IleLysLysSerValL 635
      :: :: :: :: :: :: :: :: :: :: :: :: ::
```

```
2212 TTTCTACCCATTAGACAGAGATCACCTCTGGTGTGTATCTGCAAAAGCAT 2261
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
635 euArgGlyTyrAlaLysLeuGlyTyrGlyAsn...AsnLeuProPheTyr 650
      ||| ||| ||| ||| :: :: :: :: :: :: :: :: :: :: :: ::
2262 CTCAGGATATGCAAA...GGTTTTGGAAACAAGCGTTTACCCTTCTAT 2308
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
651 GluAsnPhThrAlaGlyGlyTyrGlySerValArgGlyTyrAspGlnSe 667
      ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2309 CAAACTTATACAGGGGTGCGATGTTGTTCAATTACGCGGTTTTGCTTATGG 2358
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
667 rSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgArgGlyGlnG 684
      ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2359 TAGCATTTGGCCCTAACGCA.....ATTATCAAGGTCAA 2393
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
684 lnThrThrLeu.....GlyGluValValGlyGlyAsnAlaLeu 696
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
2394 ATAATAAATTTAATAAGATTAAGTTCTGATGTGTTGGTGAATGCAATC 2443
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
697 AlaThrPheGlySerGluLeuIleLeuProLeuProPheLysGlyAspTr 713
      ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: ::
2444 GCTACAGCTAGCCGACAGTTAATTGTCCTCACTCCATTTGTGAGTGATA 2493
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
713 pIle...AspGlnValArgProValIlePheIleGluGlyGlyGlnValP 729
      :: :: :: :: :: :: :: :: :: :: :: :: :: ::
2494 GAGTCAAAATACAGTCGGAACCTCCCTATTATTGTTGATGCGCAAGTGTT 2543
      :: :: :: :: :: :: :: :: :: :: :: :: ::
729 heAspThrThr...GlyMetAspLysGlnThrIleAspLeuThrGlnPhe 744
      ::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2544 GGAATACTAATGGAATCAGATAAAATGGATTAGAGACCAATGTCTTG 2593
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
745 LysAspProGlnAlaThrAlaGluGlnAsnAlaLysAlaAlaAsnArgPr 761
      ||| ||| :: :: :: :: :: :: :: :: :: :: :: :: ::
2594 AAAGAC.....TTACC 2604
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
761 oLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAlaT 778
      : :: :: :: :: :: :: :: :: :: :: :: :: ::
2605 CGATTATGCAAAATCAAGCGCTACTCGCGCTCTACAGGTGTCGGAATCC 2654
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
778 hrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLeu 794
      ||| :: :: :: :: :: :: :: :: :: :: :: :: ::
2655 AATGGCAATCTCTAGTGGACCAGTGGTATTCTTATGCTAAACCAAT 2704
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
795 AsnLysLysGlnAsnAspGlnThrAspThrValGlnPheGlnIleGlySe 811
      :: :: :: :: :: :: :: :: :: :: :: :: ::
2705 AAAAAATACAAATGATGATGTCGAACACTTCCAATTTAGTATTGGGG 2754
      :: :: :: :: :: :: :: :: :: :: :: :: ::
811 rValPhe 813
      : |||
2755 TTCTTTC 2761
      ::::::::::: :: :: :: :: :: :: :: :: :: :: :: :: ::
seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-119-125A-3
seq_documentation_block:
; Sequence 3, Application US/08119125A
; Patent No. 5610011
; GENERAL INFORMATION:
; APPLICANT: SMITH, Hilda Elizabeth
; APPLICANT: VECCH, Uri
; TITLE OF INVENTION: DNA Sequences which code for Virulence
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, po
; TITLE OF INVENTION: Antibodies derived therefrom and the use thereof for the di
; TITLE OF INVENTION: protection against infection by S. suis in mammals, includi
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centraal Diergeneeskundig Instituut
; STREET: Edelhertweg 15
; CITY: PH Lelystad
; STATE:
; COUNTRY: The Netherlands
; ZIP: NL-8219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
```

[illegible]

```

alignment_scores:
  Quality: 154.50      Length: 1048
  Ratio: 0.554        gaps: 55
  Percent Similarity: 41.698      Percent Identity: 18.989

alignment_block:
  US-09-701-711-2 x US-08-119-125A-3 ..

```



```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7286 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..6654
; OTHER INFORMATION: /product= "cytotactin"
PCT-US95-11684-1

alignment_scores:
  Quality: 150.00      Length: 884
  Ratio: 0.378        Gaps: 39
  Percent Similarity: 44.910  Percent Identity: 18.665

alignment_block:
  US-09-701-711-2 x PCT-US95-11684-1 ..
  Align seg 1/1 to: PCT-US95-11684-1 from: 1 to: 7286

5 TyrPheLysGlyPheGlnValSerAlaMetThrMetAlaValMetMetVa 21
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4570 TACCTCTCTGGACTTGCTCCGAGCATCCGACCAACCAACCATCAGTCCAC 4619
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
21 lMetSerThrHisAlaGlnAlaAlaAspPheMetAlaAsnAspIleThrI 38
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4620 AGCCACGACAGAGGCC.....TATA 4636

38 lThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuPro 54
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4637 TGCCCCCTCTGGAAACCTTAACCATTTCCGACATT...AATCCCTACGGG 4683
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
55 PheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspGlyVally 71
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4684 TTCACAGTTTCTGGATGGCATCGGAGATGCCTTTGACAGCTTCTAGT 4733
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
71 sAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnG 88
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4734 AACGGTGTGGATTCTGGAGAGCTCTGACCCGCCAGGAATTCACACTTT 4783
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
88 luGlyArgIleIleTyrGlnValThrGluArgPro..... 99
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4784 CAGGA.....ACCCAGAGGAAGCTGGAGCTTAGAGGC 4815
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
100 LeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGl 116
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4816 CTCATACTGGCATTTGGCTATGAG.....GTTATGGTCTCTGG 4853
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
116 yLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGlyGlnProLeuL 133
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4854 CTTACCCCAAGGGGCATCAACCC.....AAGCCCTTGA 4885
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
133 ysGlnAlaThrValGlnMetIleGluThrGluLeuThrAsnGlnTyrIle 149
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4886 GGGCTGAGATTGTTACAGAAGCCGAACCGGAAGTTGACAACCTTCTGTT 4935
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
150 SerGlnGlyTyr.....TyrAsnThrGluIleThrValLysGlnTh 163
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4936 TCAGATGCCACCCAGACGGTTTCGCTCTGCTGGACAGCTGTGAAGG 4985
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
163 rMetLeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGlyL 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
4986 GGTCTTCGACAAATTTTGTCTCAAAATCAGAGATACCAAAAGACAGTCTG 5035
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
180 ysProAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPheSer 196
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
5036 AGCCA.....CTGGAATAACCTTACTTGCCCGGACGAGTACCAGG 5076
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
197 AspAlaAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsnPr 213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

seq_name: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq:PCT-US95-11684-1

seq_documentation_block:
; Sequence 1, Application iPC/TUS9511684
; GENERAL INFORMATION:
; APPLICANT: THE SCRIPTS RESEARCH INSTITUTE
; TITLE OF INVENTION: CYTOTACTIN DERIVATIVES THAT STIMULATE
; TITLE OF INVENTION: ATTACHMENT AND NEURITE OUTGROWTH, AND METHODS OF MAKING
; TITLE OF INVENTION: AND USING SAME
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Scripps Research Institute, Office of
; ADDRESSEE: Patent Counsel
; STREET: 10666 North Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11684
; FILING DATE: 14-SEP-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/308,359
; FILING DATE: 16-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: BEC0019P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 1:

```

5077 GAC.....ATAACAGG 5087
 213 oLeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuG 230
 ||| :||| :||| ||| |||
 5088 TCTCAGAGAGGCTACTGAATACGAAATGAA...CTCTATGGAATAAGCA 5134
 230 luAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGluIle 246
 :||| :||| :||| :||| :|||
 5135 AAGGAAGCGCATCCACAGACTCAGTGCT..... 5163
 247 LysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheValG 263
 ||| :||| :||| :||| :|||
 5164 ..ATAGCAACAACAGCATCGGCTCCCAAGGAAGTCATTTTCTCAGA 5210
 263 uIleSerLeuHisGluGly..... 269
 :||| :||| :||| :||| :|||
 5211 CATCACTGAAATTCGGCTACTGTCTGAGTGGAGGACCCACGCGCCCAAG 5260
 270 .GluGlnTyrArgPheGlyGlnThrGlnPheLeuGlyAsnLeu..... 283
 ||| :||| :||| :||| :|||
 5261 TGGAGAGCTTCGGATACCTATGTGCCCATTCACGAGAGGTACACCCCTCC 5310
 284ThrTyrThrGlnAlaGluLeu..... 290
 ||| :||| :||| :||| :|||
 5311 ATGCTAACTGTGGACGGAACCAAGACTCAGACCGAGGCTGGTGAACATCAT 5360
 291GluAlaLeuLysPheLysAlaGluGluGlyPheSerG 304
 ||| :||| :||| :||| :|||
 5361 ACCTGGCGTGAGTACCTGTTCACATCATCGCCATGAAGGGCTTTGAGG 5410
 304 lnAlaMetLeuGluGlnThrThrAsnAsnIleSerThrLysPhe...Gly 319
 :||| :||| :||| :||| :|||
 5411 AAAGT.....GAACCTGTCTCAGGGTCATTCACACAGACTCTGGATGGC 5454
 320 AspAspGlyTyrTyrTyrAlaGlnIle.....ArgProValThrAr 333
 :||| :||| :||| :||| :|||
 5455 CCATCTGGCCCTGTGTGACGACCAACATCACTGACTCAGCAAGCCTTGGCCAG 5504
 333 qIleAsnAspGluSerArgThrValAspValGluTyrTyrIleAspProv 350
 :||| :||| :||| :||| :|||
 5505 GTGGCACCACCCATTGCCACTGTGGAC..... 5532
 350 alHisProValTyrValArgIleAsnPheThrGlyAsnPheLysThr 366
 :||| :||| :||| :||| :|||
 5533AGTTATGTC.....ATCTCCTACACAGC..... 5556
 367 GlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAl 383
 :||| :||| :||| :||| :|||
 5557GAGAAAGTCCACAGAAATTCACGACGCGGTGTC 5588
 383 aSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyP 400
 :||| :||| :||| :||| :|||
 5589 CGGGAACACA..... 5598
 400 hePheLysHisValThrValAspThrArgProValProAsnSerProAsp 416
 :||| :||| :||| :||| :|||
 5599 ..GTGGAGTAGTCTCTACCGCACCTCGAGCTGCC.....ACG 5634
 417 GlnValAspValAsnPheValValGluGluGlnProSerGlySerSerTh 433
 :||| :||| :||| :||| :|||
 5635 GAATPACACACTGAGAATCTTTGCAGAGAAAGGGCCCAAGAGAGCTCAAC 5684
 433 rIleAlaAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheaspv 450
 ||| :||| :||| :||| :|||
 5685 CATCACTGCCAAGTTCACACAGAC..... 5709
 450 alSerGlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPhe 466
 :||| :||| :||| :||| :|||
 5710CTCGATTCTCCAAGAGACTTGACTGCTACTGAG 5742
 467 SerArgSerGluThr..... 471
 :||| :||| :||| :||| :|||
 5743 GTTCAGTCGGAACACTGCCCTTACCTGTGGCGACCCCGGGCATCAGT 5792

472ArgGluV 474
 5793 CACCGGTACCTGCTGGTCTATGAATCAGTGGATGGCACAGCTCAAGGAAG 5842
 474 al.....TyrSerLeuGlyMetThrAsnPro 482
 ||| :||| :||| :||| :|||
 5843 TCATGTGGTCCAGATACCACTCTCTACAGCTGGCAGACCTGAGCCCA 5892
 483 TyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLy 499
 :||| :||| :||| :||| :|||
 5893 TCCACCCACTACAGCCCAAGATCCAGCACTCAATGGGCCCTGAGG... 5940
 499 sThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrG 516
 :||| :||| :||| :||| :|||
 5941AGCAATATGATCCAGACCATCTTCACCAACAATTG 5974
 516 lyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSer 532
 ||| :||| :||| :||| :|||
 5975 GA.....CTCCTGTACCCCTTCCCAAGGACTGCTCCCAAGCAATG... 6015
 533 PheGlyLeuAsnAlaAspAsnThrLys.....LeuHi 543
 ||| :||| :||| :||| :|||
 6016CTGAATGGAGACACGACCTCTGGCCTCTACACCATTTATCTGAA 6059
 543 sGlyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAspG 560
 :||| :||| :||| :||| :|||
 6060 TGGTGATAAGGCTCAGGCGCTGGAAGTCTTCTGTGCACATGACCTCTGATG 6109
 560 lyGlyLysIleGlnVal.....AspAsnAsnGlyIleProAspPhe 573
 ||| :||| :||| :||| :|||
 6110 GGGGTGGATGGATGTGTCTCTGAGACGCAAAAACGGACGCGAGAACTTC 6159
 574 LysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyrAsnTyrSerSe 590
 :||| :||| :||| :||| :|||
 6160 TACCAAACTGGAAGGCATATGCTGCT..... 6186
 590 rLeuAspArgProValPheProThrGlnGlyMetSerHisSerValAspL 607
 6186 6186
 607 euThrValGlyPheGlyAspLysThrHisGln..... 617
 ||| :||| :||| :||| :|||
 6187GGATTTGGGACCGCAGAGAAATCTTGGCTTGGGCTGGAC 6228
 618LysValValTyrGlnGlyAsnIleTyrArgProPheIleLy 631
 ||| :||| :||| :||| :|||
 6229 AACCTGAACAAATACAGCCCGGCGAG.....TAGAGCT 6266
 631 sLysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuP 648
 :||| :||| :||| :||| :|||
 6267 CCGGTGGACCTGCGGACCATGGGACAGAGCCTTGCT..... 6306
 648 roPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySerValArgGlyTyr 664
 :||| :||| :||| :||| :|||
 6307GTCTAT 6312
 665 AspGlnSerSerLeuGlyProArgSerGlnAlaTyrLeuThrAlaArgAr 681
 ||| :||| :||| :||| :|||
 6313 GACAAGTTCAGCTGGGAGATGCCAAGACTCGCTACAAAGCTGAAGGTGGA 6362
 681 gGlyGlnGlnThrThrLeuGlyGluValVal.....GlyGlyAsnA 695
 :||| :||| :||| :||| :|||
 6363 GGGGTACAGTGGACAGCAGGTGACTCATGGCTTACCACAAATGGCAGAT 6412
 695 laLeuAlaThrPheGlySerGluLeu..IleLeuProLeuProPheLysGl 711
 :||| :||| :||| :||| :|||
 6413 CCTTCTCCACCTTTGACAAGGACACAGATTTCAGCCATCACCA..... 6454
 711 yAspTrpIleAspGlnValArgProValIlePheIleGluGlyGlnV 728
 :||| :||| :||| :||| :|||
 6455ACTGTGCTCTGTCTACAAGGGGCTTCTG 6482


```

1111
3047 TACGGTTGAAAGTAACGCTAATTTCAAGAGCTATCACAAATTCACATTTTA 3096
808 lnileGlySerValPhe 813
3097 ATGTAGCGGGCTGTTT 3113
seq_name: /cgn2.6/ptodata/2/lna/5A_comb.seq:US-08-302-832-1

seq documentation_block:
; Sequence 1, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,832
; FILING DATE: 16-SEP-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US pct/us93/02166
; FILING DATE: 16-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstreser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-404
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-832-1

alignment_scores:
Quality: 147.00 Length: 939
Ratio: 0.347 Gaps: 53
Percent Similarity: 45.154 Percent Identity: 20.554

alignment_block:
US-09-701-711-2 x US-08-302-832-1 ...
Align seg 1/1 to: US-08-302-832-1 from: 1 to: 5116
27 GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnAr 43
681 GAAATGGTGCAGTGTTCACAAAGAAACAACTCCGCGCTATTCACCG 730
43 gValThr.....IleGluSerLeuGlnSerValLeuProPheArgL 57
|||||
```


318 heGlyAsp.....AspGlyTyrTyrTyrAlaGlnIleArgPro 330
 :|||||
 1528 GGGCGATATTGGTTAAATGACGGCAATTAACCGCTCAA..... 1568
 :|||||
 331 valThrArgIleAsnAspGluSerArgThr..... 340
 :|||||
 1569GGTAGTGGTGATATCGCTAAACCGGTTGTTGTGGAGACGTC 1612
 :|||||
 341ValAspVal.....GluT 345
 :|||||
 1613 GGGCGATATTATTTCATCAAGACAATGCAATTTGTGACGCCAAGAGT 1662
 :|||||
 345 yTyrIleAspPro.....ValHisProValTyrValArgArg 357
 :|||||
 1663 GGTGTGTACCCGGATATGATCTATTAAATGCAGAAACAGCAGGACGC 1712
 :|||||
 358 IleAsn.....PheThrGlyAsnPheLysThrGlnAs 368
 :|||||
 1713 AGCAATACTTCAGAACAGCATGAATACACGGGATCCGGGAATAGTGCCAG 1762
 :|||||
 368 pGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerA 385
 :|||||
 1763 CACCCCAAAACGAACANA.....GAAAGACACAACTTAACAA 1800
 :|||||
 385 snGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePhe 401
 :|||||
 1801 ACACAACTCTTGAG.....AGTATACTAAAAAAGGTACCTTT 1838
 :|||||
 402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418
 :|||||
 1839 GTTAACATCACTGCTAATCAACGCATCTATGTCATAGC..... 1877
 :|||||
 418 laspValAsnPheValValGluGlnProSerGlySerSerThrIleA 435
 :|||||
 1878 .TCCATTAATTA.....TCCAAATGGCAGCTTAACCTTT 1911
 :|||||
 435 laAlaGlyTyrSerGlnSerGlyValThrPheGlnPheAspValSer 451
 :|||||
 1912 GGAGTGGGCGGCGGTGGCGGTGAGATT..... 1946
 :|||||
 452 GlnAsnAsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerAr 468
 :|||||
 1947 ..AACACGATATTACCACCGGTGATGAT.....ACCAG 1978
 :|||||
 468 gSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheT 485
 :|||||
 1979 AGTGCAAACTTAACAAATTTACTACGGCGGTGGTGTGATGTTCAFAAAA 2028
 :|||||
 485 hrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLys 501
 :|||||
 2029 ATATCTCACTCGGGCGCAAGGTAAACATAACATTACAGCTAAACAAGAT 2078
 :|||||
 502 TyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySe 518
 :|||||
 2079 ATCGCCTTTGAGAAAGGAAGCAACCAAGTCATTACAGGTCAA...GGGAC 2125
 :|||||
 518 rLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyL 535
 :|||||
 2126 TATTAACCTCAGGC.....AATCAAAA.....GGTT 2151
 :|||||
 535 euAsnAlaAspAsnThrLysLysHisGly..... 544
 :|||||
 2152 TTAGATTAAATAATGCTCTCTAAACGGCAGCTGGCAGCGACTGCAATTC 2201
 :|||||
 545GlyArgPheMetGlyI 550
 :|||||
 2202 ACCACTAAAGAAACCAATAAATACGCTATCACAAATAAATTTGAAGGGAC 2251
 :|||||
 550 eSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspA 567
 :|||||
 2252 TTAAATATT.....TCAGGAAAGTGAACATC...T 2280

567 snAsnGlyIleProAspPheLysHisAspTyrThrThrTyrAsnAlaIle 583
 :|||||
 2281 CAATGGTTTTACCTAAATGAAAGTGGATATGATAAATTCAAAGGACGC 2330
 :|||||
 584 LeuGlyTyrAsnTyrSerSerLeuAsp..... 592
 :|||||
 2331 ACTTACTGGAATTTAACTCTTAAATGTTCCGAGAGTGGCGAGTTAA 2380
 :|||||
 593ArgP 594
 :|||||
 2381 CTTCACTATTGACTCCAGAGGAAGCATAGTCAGGCACACTTACCCAGC 2430
 :|||||
 594 roValPheProThrGlnGlyMetSerHisSerValAspLeuThrValGly 610
 :|||||
 2431 CT...TATAATTTAAACGGTATATCATTCACAAACACACTACC..... 2471
 :|||||
 611 PheGlyAspLysThrHisGlnLysValValTyrGlnGlyAsnIleTyrAr 627
 :|||||
 2472 TTTAATGTTGACGAAATGCAAGAGTCAACTTTGAC..... 2507
 :|||||
 627 gProPheIleLysLysSerVal...LeuArgGlyTyrAlaLysLeuGlyT 643
 :|||||
 2508ATCAAGGCACCAATAGGATAAATAAGTATTCTAGTTGAATT 2550
 :|||||
 643 yTyrGlyAsnAsnLeuProPheTyrGluAsnPheTyrAlaGlyGlyTyrGly 659
 :|||||
 2551 ACGCATCA.....TTTAAATGGAACATTTTCAGTTTCGGGAGGGGG 2591
 :|||||
 660 SerValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAla... 675
 :|||||
 2592 AGGTGTT.....GATTTACACTTCGCGCTCATCTCTAAGCTCA 2632
 :|||||
 676TyrLeuThrAlaArgGlyG 683
 :|||||
 2633 ACCCCCGGTGAGTTATAAATCTTAATCTTAATGTTTCAACAGGGT 2682
 :|||||
 683 lngln.....ThrThrLeuGlyGluValValGlyGlyAsnAlaLeu 696
 :|||||
 2683 CAAGTTTAAGATTAAAACTTCAGGCTCAACAAAACTGCGTTCCTCAATA 2732
 :|||||
 697AlaThrPheGlySerGluLeuLeuLeuPr 706
 :|||||
 2733 GAGAAAGATTTAACTTTAAATGCCACGGGCAACATAACACTTTTGCA 2782
 :|||||
 706 oLeuProPheLysGlyAspTyrIleAspGln.....ValArgProV 720
 :|||||
 2783 AGTTGAGGCACCGATGGAATGATGTTAAAGGCATTTAGCCAAAAAAA 2832
 :|||||
 720 alilePheIleGlyGlyGlnValPheAspThrThrGlyMetAspLys 736
 :|||||
 2833 ACATAACCTTTGAAGGAGGTAAACATC.....ACCTTTGGCTCCAGAAA 2876
 :|||||
 737 GlnThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluG 753
 :|||||
 2877 GCGTAAACAGAAATC.....GAAGCAATGTTACTATCAATAA 2914
 :|||||
 753 nasnAlaLysAlaAla.....AsnArgP 761
 :|||||
 2915 CAACGCTAACGTCACCTTTATCGGTTCCGATTTTGACAAACCATCAAAAAC 2964
 :|||||
 761 roLeuLeuThrGlnAspLysGlnLeuArgTyrSerAlaGlyValGlyAla 777
 :|||||
 2965 CTTTAACATTAAAAAAGATGTCATCATTAATAGCGCAACCTTACCGCT 3014
 :|||||
 778 ThrTrpTyrThrProIleGlyProLeuSerIleSerTyrAlaLysProLe 794
 :|||||
 3015GGAGCAATATGTCATATATAGCCGGAATCT 3046
 :|||||
 794 uAsnLysLysGlnAsnAspGlnThrAspThrVal.....GlnPheG 808
 :|||||
 3047 TACCGTTGAAGTAACGCTAATTTCAAGCTATCACAAATTTTCACTTTTA 3096
 :|||||
 808 lngIleGlySerValPhe 813

341ValaspVal...GluT 345
1613 GGGCGATGATTATTATTCATCAAAGACAATGCAATTTGTTTCAGCCCAAGAGAT 1662
345 yTyRIleAspPro.....ValHisProValTyRValArgArg 357
1663 GGTGTGTAGACCCGGTAATGTATCTATTATTCGACAAACAGCAGGACGC 1712
358 lI'eAsn.....PheThrGlyAsnPhelYsThrGlnAs 368
1713 AGCAATACTTCAGAAGACGATGAATACACGGATCCGGGAATAGTGCCAG 1762
368 pGluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerA 385
1763 CACCCEAAAGCAACAAA.....GAAAGACACACATTAAACA 1800
385 snGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePhe 401
1801 ACACAACTCTTGAG.....AGTATCTTAAAAAAGGTACCTTT 1838
402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418
1839 GTTAACATCACTGCTAATCAACGCATCTATGCTCAATAGC..... 1877
418 lAspValAsnPhelValValGluGlnProSerGlySerSerThrIleA 435
1878 .TCCATTAAATTA.....TCCATGGCAGCTTAACCTTTT 1911
435 laAlaGlyTyRSerGlnSerGlyGlyValThrPheGlnPheAspValSer 451
1912 GGAGTGAGGTGGAGCGGTGGCGCGCTTGAGATT..... 1946
452 GlnAsnAsnPhelMetGlyThrGlyLysHisValAsnAlaSerPheSerAr 468
1947AACACGATATTACCACCGGTGATGAT.....ACCAG 1978
468 gSerGluThrArgGluValTyRSerLeuGlyMetThrAsnProTyRPhet 485
1979 AGGTGCACAAATTACAATTTTACTCAGCGCGCTGGGTGTGTTTCATAAAA 2028
485 hrValAsnGlyValSerGlnSerLeuSerGlyTyTyRArgLysThrLys 501
2029 ATATCTCACTCGGCGCGCAAGGTAAACATAACATTACAGCTAAACAAGAT 2078
502 TyrAspAsnLysAsnIleSerAsnTyRValLeuAspSerTyRgLyGlyse 518
2079 ATCGCGCTTGAGAAGGAAGCAACCAAGTCATTACAGGTCAA...GGAC 2125
518 rLeuSerTyRgLyTyRProIleAspGluAsnGlnArgIleSerPheGlyL 535
2126 TATTACTTCAGGC.....AATCAAAA.....GGTT 2151
535 euAsnAlaAspAsnThrLysLeuHisGly..... 544
2152 TTAGATTAAATGTCCTCTCTAAACGGCACTGGCAGCGACTGCAATTC 2201
545GlyArgPheMetGlyI 550
2202 ACCACTAAAGCAACCAATAATACGCTATCACAAATAAATTTGAAGGAC 2251
550 eSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspA 567
2252 TTTAATATT.....TCAGGGAAGTGAACATC...T 2280
567 snAsnGlyIleProAspPheLysHisAspTyRThrThrTyRAsnAlaIle 583
2281 CAATGGTTTTACCCTAAATAAGTGAAGTGGATATGATAAATCAAGGACGC 2330
584 LeuGlyTyPAsnTyRSerSerLeuAsp..... 592
2331 ACTTACTGGAATTTAAACCTCTCTTAATGTTTCCAGAGATGGCGAGTTTAA 2380
593Arqp 594

[illegible]

```
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-469-880-1-
seq_documentation_block:
; Sequence 1, Application US/08469880
; Patent NO. 5876733
; GENERAL INFORMATION:
```

APPLICANT: Barenkamp, Stephen J.
TITLE OF INVENTION: High Molecular Weight Surface Proteins
TITLE OF INVENTION: of No. 5876733-Typeable Haemophilus
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd.
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/469,880
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIORITY APPLICATION DATA: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIORITY APPLICATION DATA: US 08/302,832
FILING DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5116 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-1
alignment_scores:
Quality: 147.00 Length: 939
Ratio: 0.347 Gaps: 53
Percent Similarity: 45.154 Percent Identity: 20.554
alignment_block:
US-09-701-711-2 x US-08-469-880-1 ..
Align seg 1/1 to: US-08-469-880-1 from: 1 to: 5116
27 GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnAr 43
::: :::::||||| ::::: ::::: ::::: ::::: |||
681 GAAATGGTGCAGTTTTCACAGAAAGAAACAACTCCGCCGTATTCACCG 730
43 gValThrIleGluSerLeuGlnSerValLeuProPheArgL 57
||||| ||:::||||| ::::: ::::: ::::: |||
731 TGTTCATCTAACCAATCTCCCAATTAAGAGGATTTTAGATCTAAC. 779
57 euGlyGlnVal.....ValSerGluAsnGlnLeuAlaAspGlyValLys 71
||||| ::::: ||| ::::: |||
780 ..GGCAAGTCTTTTATCAACCAATGGTATCAATAGGTAAGAC 827
72 AlaLeuTyraIleThrGlyAsnPheSer.....AspValGlnVa 84
||||| ::::: ||| ::::: |||
828 GCAATTATTAACTAATAGGTTTACGGCTTCTACGCTAGACATTTCTAA 877

84 lTyHisGlnGluGlyArg...IleIleTyThrGlnValThrGluArgProL 100
::: :::::||||| ::::: ::::: |||
878 CGAAACATCAAGCGCGTATTTTACCTTCGAGCAACCAAGATAAG 927
100 euIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGly 116
::: ||||| ::::: |||
928 CGCTCGCTGAAATGTGAATCAGCGTTTAAATCTGTCGGTAAAGACGCG 977
117LeuGlnGluGlyLeuLysAsnAlaGly..... 125
::: :::::||||| ::::: |||
978 AGTGTAAATCTATTATGGGCAAGTGAACAAACAGAGGTTGTCATTAGCGT 1027
126LeuAlaValGlyGlnProLeuLysGlnAlaAr 136
||| :::::||||| ::::: |||
1028 AATGGTGGCAGCATTTCTTTACTCGCAGGGCAAAATC.....A 1068
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrlleSerGlnGly 152
||::: |||
1069 CCATCAGCGATATA..... 1082
153 TyrTyAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnAr 169
||| ||||| ::::: |||
1083ATAACCCCAACCATTTACTTACAGCATTCGCCGCCCTGAAATGAAGC 1129
169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValVala 186
||::: |||
1130 GGTCAATCTGGCGATATTTTTCACAAAGCGGT.....A 1164
186 spIleAsnIle.....IleGlyAsnGlnHisPheSerAspAla 198
::: ||||| ::::: |||
1165 ACATTAATGTCGTCGTCGTCACCTATTCGAACCAAGTAACTTTCTGCT 1214
199 AspLeuIleAspValLeuAlaIleLysAspAsnLysIleAsn...ProLe 214
||| ::::: |||
1215 GATTCGTGAAGC.....AAGATAAAGCGCAATATTCTTCT 1252
214 uSerLysAlaAspArgTyThrGlnGluLysLeuValThrSerLeuGluA 231
|||| ::::: |||
1253 TTCGCCCAAGAGGGTGAAAGGAAATTCGCGGTGTAATTCGCTCAA 1302
231 snLeuArgAlaLysTyThrLeuAsnAlaGlyPheValArgPheGluIleLys 247
|| ::::: |||
1303 ATCAGCAAGCT.....AAA 1316
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIle..... 260
::: ||||| ||::: |||
1317 GCGCGCAAGCTGATGATTACGGCGATAAAGTCACATTAAACACAGTGC 1366
261 .PheValGluIleSerLeuHisGlu...GlyGluGlnTyThrArgPheGlyG 276
::: ||||| ||| ||||| |||
1367 AGTTATCGACCTTTCAGGTAAGAAGGGGAGAACTTACCTTGGCGGTG 1416
276 lnthrGlnPheLeuGlyAsnLeuThrTyThrGlnAlaGluLeuGluAla 292
::: ||| ::::: |||
1417 ACGAGCGCGCGAAGGT.....AAAAAGGCGATTCATTA 1451
293 LeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetLeuGluG 309
||| ::::: |||
1452 GCAAGAAACCTCTTTAGAAAAGGC..... 1478
309 nthrThrAsnIleSerThrLys.....P 318
||| ||||| |||
1479 .TCAACCATCAATGTATCAGCAAGAAAGGCGGCGCGCTATTGTGT 1527
318 heGlyAsp.....AspGlyTyThrTyThrAlaGlnIleArgPro 330
||| ||||| |||||
1528 GGGCGGATTTGCGTTAATTCAGCGCATATTAACGCTCAA..... 1568
331 ValThrArgIleAsnAspGluSerArgThr..... 340
||| ::::: |||
1569GGTAGTGGTGATATCGCTAAACCGGTGTTTGTGGACACGTC 1612
341ValAspVal...GluT 345

```
1613 GGGCGATGATTATTATCAAGCAATGCAATGTTGACGCCAAGAGT 1662
345 yTYrIleAspPro.....ValHisProValTyrValArgArg 357
1663 GGTGTAGACCCGGATAATGATCTAATTAATGAGAAACAGCAGGACGC 1712
358 IleAsn.....PheThrGlyAsnPhelYsThrGlnAs 368
1713 AGCAATACTTCAGAACGAGATGAATACACGGGATCCGGGAATAGTGCAG 1762
368 pGluValLeuArgGluMetArgGlnLeuGluGlyAlaLeuAlaSerA 385
1763 CACCCCAAAACAGAAACAAA.....GAAAAGACACATTAACAA 1800
385 snGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePhe 401
1801 ACACAACCTCTTGAG.....AGTATACTAAAAAAGGTACCTTT 1838
402 LysHisValThrValAspThrArgProValProAsnSerProAspGlnVa 418
1839 GTTAACATCACTGCTAATCAACGCATCTATCTCAATAGC..... 1877
418 lAspValAsnPhelValValGluGluGlnProSerGlySerSerThrIleA 435
1878 .TCCATTAAATTA.....TCCAATGGCAGCTTAACCTCTT 1911
435 laAlaGlyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSer 451
1912 GGAGTGAGGGTCGAGCGGTGGCGCGGTGGAGATT..... 1946
452 GlnAsnAsnPhelMetGlyThrGlyLysHisValAsnAlaSerPheSerAr 468
1947 ...AACACGATATTACCACCGGTGATGAT.....ACCAG 1978
468 gSerGluThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheT 485
1979 AGTGCAAACTTAACAATTTACTCAGCGCGGTGGTTGATGTTTCATAAAA 2028
485 hrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrArgLysThrLys 501
2029 ATATCTCAGTCGGCGCGAGGTAACTAACATAACATTCAGCTAACAAAGAT 2078
502 TyrAspAsnLysAsnIleSerAsnTyrValLeuAspSerTyrGlyGlyse 518
2079 ATCGCCTTTGAGAAAGGAGCAACCAAGTCATTACAGGTCAA...GGGAC 2125
518 rLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIleSerPheGlyL 535
2126 TATTACCTCAGGC.....AATCAAAA.....GGTT 2151
535 euAsnAlaAspAsnThrLysLeuHisGly..... 544
2152 TTAGATTAAATATGTTCTCTCTAAACGGCAGCTCGCAGCGGACTCAATTC 2201
545 .....GlyArgPheMetGlyI 550
2202 ACCACTAAAGAACCAATAATACGCTATCACAAATAAATTTGAAGGAC 2251
550 eSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIleGlnValAspA 567
2252 TTTAAATATT.....TCAGGGAAGTGAACATC...T 2280
567 snAsnGlyIleProAspPhelYsHisAspTyrThrTyrAsnAlaIle 583
2281 CAATGGTTTTACCTACCAAAATCAAGTGGATATGATAAATTCAAAGGACGC 2330
584 LeuGlyTrpAsnTyrSerLeuAsp..... 592
2331 ACITACTGGAATTAACCTCTCTAAATGTTTCCGAGAGTGGCGAGTTAA 2380
593 .....Argp 594
```

```
seq_name: /cn2_6/ptodata/2/ina/5B_COMB.seq: US-08-728-470-1
seq_documentation_block:
; Sequence 1, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
```

Thu Sep 19 10:04:26 2002

```

;
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5116 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-728-470-1

alignment_scores:
  Quality: 147.00      Length: 939
  Ratio: 0.347         Gaps: 53
  Percent Similarity: 45.154      Percent Identity: 20.554

alignment_block:
US-09-701-711-2 x US-08-728-470-1 ..

Align seg 1/1 to: US-08-728-470-1 from: 1 to: 5116

27 GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGluGlnAr 43
   :::::::::::::::::::::
681 GAAATGGTGCAGTCTTTTACAGAAACAAACAACTCCGCGTATTCAACG 730
   :::::::::::::::::::::
43 gValThr.....IleGluSerLeuGlnSerValLeuPropheArgL 57
   |||||
731 TGTTCATCTAACCAAAATCTCCCAATTAAGGGATTTTAGATTCTAAC. 779
   |||||
57 euGlyGlnVal.....ValSerGluAsnGlnLeuAlaAspGlyVallys 71
   |||||
780 ..GGACAAGTCTTTTAAATCAACCAAGGATGATACATAAGTAAAGAC 827
   |||||
72 AlaLeuTyraLThrGlyAsnPheSer.....AspValGlnVa 84
   |||||
828 GCAATTATTAACTAATAGCTTTTACGGCTTCTACGCTAGACATTCTAA 877
   |||||
84 lTyRHIsGlnGlyArg...IleIleTyRGlnValThrGluArgProL 100

```

```

878 CGAAACATCAAGGCGCGTAATTTACCTTCAGCAACCAAGATAAAG 927
   :::::::::::::::::::::
100 euIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGly 116
   :::::::::::::::::::::
928 CGCTCGCTGAAATTTGTGAATCAACGCGTTTAACTACTGTCGCTAAGACGGC 977
   :::::::::::::::::::::
117 .....LeuGlnGluGlyLeuLysAsnAlaGly..... 125
   :::::::::::::::::::::
978 AGTGTAATCTTTATTTGGTGCAGCAAGTGAAACAGGAGGTGTGATTAGCGT 1027
   :::::::::::::::::::::
126 .....LeuAlaValGlyGlnProLeuLysGlnAlaL 136
   |||||
1028 AAATGGTGCAGCATTTCTTTACTCGAGGCAAAATC.....A 1068
   |||||
136 hrValGlnMetIleGluThrGluLeuThrAsnGlnTyrlleSerGlnGly 152
   |||||
1069 CCATCAGCGATATA..... 1082
   |||||
153 TyrTyraAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnAr 169
   |||||
1083 .....ATRAACCAACCATTTACTTACAGCATTTCCGCGCTGAAATGAAGC 1129
   |||||
169 gValLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValVala 186
   |||||
1130 GGTCATCTGGCGCATATTTTCCCAAGGCGGT.....A 1164
   |||||
186 spIleAsnIle.....IleGlyAsnGlnHisPheSerAspAla 198
   |||||
1165 ACATTATGTCGTCGTCACCATTTTCGAAACCAAGGTAAATCTTCGCT 1214
   |||||
199 AspIleAspValLeuAlaIleLysAspAsnLysIleAsn....ProLe 214
   |||||
1215 GATTCTGTAAAGC.....AAAGATAAAGCGCGCATATTTGCTCT 1252
   |||||
214 uSerLysAlaAspArgTyThrGlnGluLysLeuValThrSerLeuGluA 231
   |||||
1253 TTCGCCAAGAGGGTGAAGCGGAATTTGGCGGTGTAATTTCCGCTCAA 1302
   |||||
231 snLeuArgAlaLysTyThrLeuAsnAlaGlyPheValArgPheGluIleLys 247
   |||||
1303 ATCAGCAAGCT.....AAA 1316
   |||||
248 AspAlaLysLeuAsnIleAsnGluAspLysAsnArgIle..... 260
   |||||
1317 GCGCGCAAGCTGATGATTACAGCGCATAAAGTCACATTAAAAACAGGTGC 1366
   :::::::::::::::::::::
261 .PheValGluIleSerLeuHisGlu...GlyGluGlnTyArgPheGlyG 276
   :::::::::::::::::::::
1367 AGTTATCGACCTTTCAGGTAAGAGAGGGGAGAAAACCTTACCTTGGCGGTG 1416
   :::::::::::::::::::::
276 InThrGlnPheLeuGlyAsnLeuThrTyThrGlnAlaGluLeuGluAla 292
   :::::::::::::::::::::
1417 ACGAGCGCGCGAAGGT.....AAAAGGCATTCAATTA 1451
   |||||
293 LeuLeuLysPheLysAlaGluGlyPheSerGlnAlaMetLeuGluGl 309
   |||||
1452 GCAGAGAAAACCTCTTTAGAAAAAGGC..... 1478
   :::::::::::::::::::::
309 nThrThrAsnAsnIleSerThrLys..... 318
   :::::::::::::::::::::
1479 .TCAACCATCAATGATATCAGGCAAAAGAAAGCGGACCGCGGTATTGTGT 1527
   :::::::::::::::::::::
318 hseGlyAsp.....AspGlyTyThrTyThrAlaGlnIleArgPro 330
   |||||
1528 GGGGCGATATTCGGTTAATTGACGGCAATATTAAACGCTCAA..... 1568
   |||||
331 ValThrArgIleAsnAspGluSerArgThr..... 340
   :::::::::::::::::::::
1569 .....GGTAGTGGTATATCGTAAACCGCGGTGGTGGTGGAGACGTC 1612
   |||||
341 .....ValAspVal...Glut 345
   |||||

```



```

alignment_scores:      Length: 939
                        Quality: 147.00
                        Ratio: 0.347
                        Gaps: 53
                        Percent Identity: 20.554

alignment_block:
US-09-701-711-2 x US-08-617-697-1  ..

Align seg 1/1 to: US-08-617-697-1 from: 1 to: 5116

27  GlnAlaAspPheMetAlaAsnAspIleThrFileThrGlyLeuGlnAr 43
    ::::: ::::: |||::: ::::: ::::: ::::: |||
681 GAATGGTGCAGTTTTACAGAAACACAACTCCGCCGTATTCAACCG 730
    43 gValThr.....IleGluSerLeuGlnSerValLeuProPheArgL 57
    ||||| |||::: |||::: |||::: |||::: |||::: |||
731 TGTATACCTAACCAATCCCAATTAAGAGGATTTTAGATTCTAAC. 779
    57 euGlyGlnVal.....ValSerGluAsnGlnLeuAlaAspGlyValLys 71
    ||||| |||::: |||::: |||::: |||::: |||::: |||
780 ..GGACAAGTCTTTTAAATCAACCCAAATGGTATCATCAATAGGTAAAGAC 827
    |||::: |||::: |||::: |||::: |||::: |||::: |||
782 GCAATTATTAAACACTAATGCTTTTACGGCTTCTACGCTAGACATTCTCAA 877
    84 lTyrHisGlnGluGlyArg...IleIleTyrGlnValThrGluArgProL 100
    ::::: |||::: |||::: |||::: |||::: |||::: |||
878 CGAAAACATCAAGCGCGGTAAATTTTCACCTTCGAGCAACCAAGATAAAG 927
    100 euIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProLysGluGly 116

```



```

alignment_scores:      Length:      939
                        Quality: 147.00
                        Ratio:   0.347
                        Gaps:     53
                        Percent Identity: 20.554

alignment_block:
US-09-701-711-2 x US-08-719-641-1 ..

Align seg 1/1 to: US-08-719-641-1 from: 1 to: 5116

27  GlnAlaAlaAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnAr 43
   ;; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
681  GAAATGGTGAGTATTTTACAGAAACAAACATCCCGCGTATTCAACCG 730
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
43  gValThr.....IleGluSerLeuGlnSerValIeuProPheArgL 57
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
731  TGTACATCTAACCAATCTCCCAATTAAAGGGATTTTAGATTCTAAC. 779

57  euGlyGlnVal.....ValSerGluAsnGlnLeuAlaAspGlyVallys 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
780  ..GGACAAAGTCTTTTATCAACCCAAATGGTATCACAAATAGGTAAACAC 827
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
72  AlaleuTyraIaThrGlyAsnPheSer.....AspValGlnVa 84
   ||| : : : : : : : : : : : : : : : : : : : : : : : :
828  GCAATTATTACACTAATGGCTTTTACGGCTTCTACGCTAGACATTCTAA 877
   : : : : : : : : : : : : : : : : : : : : : : : :

84  lTyrrHisGlnGluGlyArg....IleIleTyGlnValThrGluArgProL 100
   : : : : : : : : : : : : : : : : : : : : : : : :
878  CGAAACATCATCAAGCGCGTAAATTTACCTTCGAGCAACCAAGATAAAG 927
   : : : : : : : : : : : : : : : : : : : : : : : :
100  euIleAlaGluIleAsnPheGluGlyAsnArgLeuIleProlysGluGly 116
   : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

us-09-701-711-2-sep1std.rni

Thu Sep 19 10:04:26 2002

```

; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,685A
; FILING DATE: No. 6214981ember 12, 1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Baldwin, Geraldine F.
; REGISTRATION NUMBER: 31,232
; REFERENCE/DOCKET NUMBER: 7969-060
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-968-685A-9

alignment_scores:
  Quality: 146.50      Length: 704
  Ratio: 0.467         Gaps: 37
  Percent Similarity: 44.602      Percent Identity: 21.449

alignment_block:
  US-09-701-711-2 x US-08-968-685A-9 ..

Align seg 1/1 to: US-08-968-685A-9 from: 1 to: 9542

34 AsnAspIleThrLeuGlyLeuGlnArgValThrIleGlu..... 47
   ||| |||..... |||
2648 AATGGTATACACTTGCAGCGCAACACTTACATCACCACCAAGCAAGTTGG 2697
   ||| |||..... |||
48 .....SerLeuGlnSerValLeuProPhe..... 55
   ||| |||..... |||
2698 CTTTGCTAAGCAAGATGGTTTCACTGTGATAAAGCAACCTTCTTGATA 2747
   ||| |||..... |||
56 .....ArgLeuGlyGlnVal...ValSerGluAsnGlnLeuAla 67
   ||| |||..... |||
2748 AGGACAGCTAAAGTGGGTGAAGTTGAGATTACCACCAAGCGCATTAAT 2797
   ||| |||..... |||
68 AspGlyValIleAlaLeuThrAlaThrGlyAsnPheSerValGlnVa 84
   ||| |||..... |||
2798 GCAGGTGTAAGCATCAGAGACTACAGCAATACCTTAACCGATGCCAC 2847
   ||| |||..... |||
84 lTyHisGlnGluGlyArgile.....IleTyGlnValThrG 97
   ||| |||..... |||
2848 CAAGCAACACAGCGCATGTAACTCAATTCGGTATCGTTGATAGTACTG 2897
   ||| |||..... |||
97 luArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIlePro 113
   ||| |||..... |||
2898 ACAAAACCGTCCGCCAGCATGGT..... 2923
   ||| |||..... |||
114 LysGluGlyLeuGlnGluGly.....LeuLysAsnAlaGlyLeuAlaVa 128
   ||| |||..... |||
2924 ...GATGGCTAAACGAGCGCTTTAACTTAAATAATATGTTGAGCCCAA 2970
   ||| |||..... |||
128 lGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuT 145
   ||| |||..... |||
2971 AGACTTGTCTCCACTTATGACACTGTGATTTTATCAATGGAATGCCA 3020

```

```

145 hrAsnGlnTyrIle.....SerGlnGlyTyrTyrAsn 155
   ||| |||..... |||
3021 CCACCGCTAAAGTCACTTATGATGGCAAGCCAGTAAAGTGGCGTATGAT 3070
   ||| |||..... |||
156 ThrGluIle.....ThrValLysGlnThrMetLeuAspGlyAsnAr 169
   ||| |||..... |||
3071 GTCATGTGGATGTACAAACCATTCATCAACAGCGCGCTGATGCAATAA 3120
   ||| |||..... |||
169 g.....ValLysLeuAspMetThrPheAlaGluG 179
   ||| |||..... |||
3121 AAACCAAAATTCGCGTAAACCAACACACACATCACCACCAAGCATGCTAAAG 3170
   ||| |||..... |||
179 lLysProAlaArgValValAspIleAsnIleIleGlyAsnGlnHisPhe 195
   ||| |||..... |||
3171 GTGATAAAGCA.....ATTAACTTTAGTGTAACTCTGCT... 3205
   ||| |||..... |||
196 SerAspAlaAspLeuIleAspValLeuAlaIleLysAspAsnLysIleAs 212
   ||| |||..... |||
3206 GATGACAAAGCCCTTATTAAACGCCAAAGACATCGCCGCAAT...CTAAA 3252
   ||| |||..... |||
212 nProLeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSerL 229
   ||| |||..... |||
3253 CACCCTAGCTGTGTAATTCGCAACCAAGGACGACGACGACGCGCC 3302
   ||| |||..... |||
229 euGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPheGlu 245
   ||| |||..... |||
3303 TACAAACC.....TTTCAA 3316
   ||| |||..... |||
246 lLeLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlePheVa 262
   ||| |||..... |||
3317 GTCAAAAAGTCAAGAAAAATGCTGATGATGATGATGATGATGATGATGAT 3366
   ||| |||..... |||
262 lGluIleSerLeuHisGluGluGlnTyrArgPheGlyGlnThrGlnP 279
   ||| |||..... |||
3367 CACCGTG.....GGTAAAGATGCAAAACCAACCAATCAAGTCAACA 3404
   ||| |||..... |||
279 heLeu..... 280
   ||| |||..... |||
3405 CCCTAAACCTCAAGGTAAACCGCTTGTGATATTCAACCAATAAAGAT 3454
   ||| |||..... |||
281 GlyAsnLeuThrTyr.....ThrGlnAlaGluLeuGluAla..... 292
   ||| |||..... |||
3455 GGTAACGGTTACCTTGGCTTAACACCCAAAGCGGTCTTAAAGCCGCGCAA 3504
   ||| |||..... |||
293 .....LeuLeuLysPheLysAlaGluGluGlyPheSerGlnAlaMetL 307
   ||| |||..... |||
3505 CAACACCACTCTAAC.....AACATGGCTTGTCT.....A 3536
   ||| |||..... |||
307 euGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspGlyTyr 323
   ||| |||..... |||
3537 TTAACCAACCGCTGGTAAACGACAAATCCAAAGTCGGTGTGATGGCGTG 3586
   ||| |||..... |||
324 TyrTyrAlaGlnIleArg.....ProVa 331
   ||| |||..... |||
3587 AAGTTTCCCAAGGTAAATAATGTTGTAGTGGCTGGCTGGCTGGCTGGCTG 3636
   ||| |||..... |||
331 lThrArgIleAsnAspGluSerArgThrValAspValGluTyrTyrIleA 348
   ||| |||..... |||
3637 AACTCGCAT..... 3646
   ||| |||..... |||
348 spProValHisProValTyrValArgArgIleAsnPheThrGlyAsnPhe 364
   ||| |||..... |||
3646 ..... 3646
   ||| |||..... |||
365 LysThrGlnAspGluValLeuArgArgGluMetArgGlnLeuGluGlyAl 381
   ||| |||..... |||
3647 ...ACCAGAGATGAAT.....GGCTT 3666
   ||| |||..... |||
381 aLeuAlaSerAsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgT 398
   ||| |||..... |||
3667 TGCTGGGACTAATGCTCACTTGATTAAGCAAAACCCACCTAAGCAAAAG 3716
   ||| |||..... |||

```

```

398 hrGlyPhe.....PheLysHisValThrValAspThrArgPro 410
    |||:::
3717 ACGGCATTACCGCAGGTGTAAAGAGATTACCAACATTCATCAGGTGAG 3766
    |||:::
411 Val....ProAsnSerProAspGlnVal.....AspVa 420
    :::: ||||| ||| |||
3767 ATGGCCAAACACGCAATGATGCTGTGACAGCGGCGCAAGATTATGATTT 3816
    :::: ||||| ||| |||
420 lAsnPheValValGluGluGlnProSerGlySerSerThrIleAlaAlaG 437
    :::: ||||| ||||| ||||| |||||
3817 AAAACCCAGCACTTGAACAAATCAGCTACTGCCAAACAGCACAAA 3866
    :::: ||||| ||||| ||||| |||||
437 lyTyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAsn 453
    :::: ||||| ||||| ||||| |||||
3867 ACTCATTACACGAATTCCTCAGTAGCA.....GATCAACAAGGTAAT 3907
    :::: ||||| ||||| ||||| |||||
454 AsnPheMetGlyThrGlyLysHisValAsnAlaSerPheSerArgSerG 470
    ||||| ||| :::: ||||| ||||| |||||
3908 AACTTT.....ACGGTTAGTAACCTTACTCCAGTTATGACACCTCAA 3951
    ||||| ||||| ||||| ||||| |||||
470 uThrArgGluValTyrSerLeuGlyMetThrAsnProTyrPheThr...V 486
    ||||| ||||| ||||| ||||| |||||
3952 GACCTCTGATGTCATCACCCTTGCAGGTGAAACAGGCATACACCAAGC 4001
    ||||| ||||| ||||| ||||| |||||
486 alAsn.....GlyValSerGlnSerLeuSerGlyTyr 496
    ||||| ||||| ||||| ||||| |||||
4002 TAAATAAAGGTGTGGTGCCTGTGGGCATTGACCAACCAAGGCTTAACC 4051
    ||||| ||||| ||||| ||||| |||||
497 TyrArgLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAs 513
    ||||| ||||| ||||| ||||| |||||
4052 ACGCCTAAGCTACCGTGGGTAAATAATGCAAGGCAATGTCATTGA 4101
    ||||| ||||| ||||| ||||| |||||
513 pSerTyrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnA 530
    ||||| ||||| ||||| ||||| |||||
4102 CAGCCAA..... 4108
530 rgIleSerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyArg 546
    ||||| ||||| ||||| ||||| |||||
4109 .....AATGTCAAAATACC..... 4123
547 PheMetGlyIleSerAsn...ValLysGlnLeuMetAlaAspGlyGly 562
    :::: ||||| ||||| ||||| ||||| |||||
4124 ATCAGAGGACTAAGCAACACTCTAGCTAATGTTACCAATGATAAAGGTAG 4173
    :::: ||||| ||||| ||||| ||||| |||||
562 stIleGlnValAspAsnAsnGlyIleProAspPheLysHisAspTyrThrT 579
    :::: ||||| ||||| ||||| ||||| |||||
4174 CGTACGCCACACAGACGGCAAGATAATCAAGACGAAGACAAACCC 4223
    :::: ||||| ||||| ||||| ||||| |||||
579 hrTyrAsnAlaIleLeu.....GlyTrpAsnTyrSerSer 590
    :::: ||||| ||||| ||||| ||||| |||||
4224 GTGCCGCCAGCATTTGTGATGTCTAAGCGCAGGCTTTAAC..... 4264
591 LeuAspArgProValPheProThrGlnGlyMetSerHisSerValAspLe 607
    ||||| ||||| ||||| ||||| |||||
4265 .....TTGCAAGGCAATGGTGAAGCGGTTGACTT 4293
    ||||| ||||| ||||| ||||| |||||
607 u.....ThrValGlyPheGlyAsp..... 613
    ||||| ||||| ||||| ||||| |||||
4294 TGTCTCCACTTATGACACTGTCACTTTCGCCGATGGCAATGCCACACCG 4343
    ||||| ||||| ||||| ||||| |||||
614 .....LysThrHisGlnLysValValTyr 621
    ||||| ||||| ||||| ||||| |||||
4344 CTAAGGTGACCTATGATGACACACAGCAAAACC...AGTAAAGTGTCTAT 4390
    ||||| ||||| ||||| ||||| |||||
622 GlnGlyAsnIle 625
    :::: ||||| ||||| ||||| |||||
4391 GATGTCAAATGTG 4402
    :::: ||||| ||||| ||||| |||||

```

seq_name: /cqn2_6/ptodata/2/ina/6B_COMB.seq.us-09-627-376-3

seq_documentation_block:

; Sequence 3, Application US/09627376

; Patent No. 6342385

```

; GENERAL INFORMATION:
; APPLICANT: Qi, Fengxia      Caulfield, Page      Chen, Ping
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
; FILE REFERENCE: URB-17402/22
; CURRENT APPLICATION NUMBER: US/09/627,376
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 15567
; TYPE: DNA
; ORGANISM: Streptococcus mutans
; US-09-627-376-3

```

```

alignment_scores:
  Quality: 145.00      Length: 997
  Ratio: 0.355        Gaps: 45
  Percent Similarity: 40.923      Percent Identity: 17.452

```

alignment_block:

US-09-701-711-2 x US-09-627-376-3 ..

Align seg 1/1 to: US-09-627-376-3 from: 1 to: 15567

```

53 LeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuAlaAspG 69
    ||| |||||::: ::::: |||::: |||
2151 TTGACAGTCATGTTAGGAGAGAGATGAAGATTTCATTTCAAGATTA 2200
    ||||| ||||| ||||| ||||| |||||
69 yValLysAlaLeuTyrAlaThrGlyAsnPheSerAsp..... 81
    ||||| ||||| ||||| ||||| |||||
2201 TTTTATGTACAGAAACCACTTAGGCAACTTCTCTAATTTTCTTAGTATA 2250
    ||||| ||||| ||||| ||||| |||||
82 .....ValGlnValTyrHisGlnGluGlyArgIleIle 92
    :::: ||||| ||||| ||||| |||||
2251 CTGATATGATGGATCCTATTGAATTATTACATAATGAT..... 2288
    ||||| ||||| ||||| ||||| |||||
93 TyrGlnValThrGluArgProLeuIleAlaGluIleAsnPheGluGlyAs 109
    ||||| ||||| ||||| ||||| |||||
2289 .....CCGATATTTCGTGAA..... 2303
    ||||| ||||| ||||| ||||| |||||
109 nArgLeuIleProLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyL 126
    ||||| ||||| ||||| ||||| |||||
2304 .....GGG 2307
    ||||| ||||| ||||| ||||| |||||
126 euAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleGluThr 142
    :::: ||||| ||||| ||||| ||||| |||||
2308 TATATTGGCTTCCCATCTCTTAGATCATCTATAATAAATTAGAGAAT 2357
    ||||| ||||| ||||| ||||| |||||
143 GluLeuThrAsn.....G 147
    :::: ||||| ||||| ||||| ||||| |||||
2358 CAGATTGCAAGTACTAAGGAAAAAAGAATGCAAAAGAGACTATTTTCA 2407
    ||||| ||||| ||||| ||||| |||||
147 nTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrVal..... 160
    ||||| ||||| ||||| ||||| |||||
2408 ATACATATGCCGT.....TATAACACAGAGATCAACTCCGTTTGGCTGT 2451
    ||||| ||||| ||||| ||||| |||||
161 .....LysGlnThr 163
    ||||| ||||| ||||| ||||| |||||
2452 TTTCTGCCATCGAATAGTGGTTTTTCGAACCCACCTAGGAAGAGAAA 2501
    ||||| ||||| ||||| ||||| |||||
164 MetLeuAspGlyAsnArgValLysLeuAspMetThrPheAla..... 177
    :::: ||||| ||||| ||||| ||||| |||||
2502 TCTTGTATTGAAAAATCTCTTAATGTTGATCTTTTGGGCTTATAAAGT 2551
    ||||| ||||| ||||| ||||| |||||
178 .....GluGlyLysProAlaArgValValAspIleAsnIle 190
    ||||| ||||| ||||| ||||| |||||
2552 AGCAGATAAACTAGAAAGTATGCCCTGAAATTTTAAATACTTTAAAGTAG 2601
    ||||| ||||| ||||| ||||| |||||
190 leGlyAsn..... 192
    :::: ||||| ||||| ||||| ||||| |||||
2602 TTGCTAATAATGCTTTGCAAAAGTCAAAATGATTTTGGCTTTTAGATACA 2651
    ||||| ||||| ||||| ||||| |||||

```

[illegible][illegible]

; FILING DATE: 16-MAR-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Berkstresser, Jerry W
 ; REGISTRATION NUMBER: 22,651
 ; REFERENCE/DOCKET NUMBER: 1038-557
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 415-0813
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4794 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-617-697-7

alignment_scores:
 Quality: 136.50 Length: 886
 Ratio: 0.324 Gaps: 50
 Percent Similarity: 47.517 Percent Identity: 19.977

alignment_block:

US-09-701-711-2 x US-08-617-697-7 ..

Align seg 1/1 to: US-08-617-697-7 from: 1 to: 4794

```

1 MetArgAsnSerTyr.....PheLysGlyPheGlnValse 12
:|||||:|||||:|||||:|||||:|||||:|||||:
274 ATCCGTAATAGCTCAATGCTATCATCAATTGGAAACAATTTAACATTTGA 323
:|||||:|||||:|||||:|||||:|||||:|||||:
12 rAlaMetThrMetAlaValMetMetValMetSerThrHisAlaGlnAla 29
:|||||:|||||:|||||:|||||:|||||:|||||:
324 CCAAAATGAATGGTGCAGTTTTCACAGAAAGACGCAACTCTGCGGTTT 373
:|||||:|||||:|||||:|||||:|||||:|||||:
29 laAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThr 45
:|||||:|||||:|||||:|||||:|||||:|||||:
374 TCAACCGTGTATCATCTGAC.....CAA 396
|||||:|||||:|||||:|||||:|||||:|||||:
46 IleGluSerLeuGlnSerValLeuProPheArgLeuGlyGlnVal..... 60
|||||:|||||:|||||:|||||:|||||:|||||:
397 ATCTCCCAATTAAGGGATTTAGATTCTAAC....GGCAAGCTCTTTT 443
|||||:|||||:|||||:|||||:|||||:|||||:
61 ValSerGluAsnGlnLeuAlaAspGlyValLysAlaLeuTyAlaThrG 77
|||||:|||||:|||||:|||||:|||||:|||||:
444 AATCAACCCAAATGGTATCACATAGTAAAGACGCAATATTATTAACACTA 493
|||||:|||||:|||||:|||||:|||||:|||||:
77 LysAsnPheSer.....AspValGlnValTyrHisGlnGluGly 89
|||||:|||||:|||||:|||||:|||||:|||||:
494 ATGGCTTTACTGCTTACGCTAGACATTTCTAACGAAACATCAAGGCG 543
|||||:|||||:|||||:|||||:|||||:|||||:
90 Arg....IleIleTyrGlnValThrGluArgProLeuIleAlaGluLeuAs 105
|||||:|||||:|||||:|||||:|||||:|||||:
544 CGTAATTTACCCCTTGAGCAACACGAGGATAAGCACTCGCTGAATCGT 593
|||||:|||||:|||||:|||||:|||||:|||||:
105 nPheGluGlyAsnArgLeuIleProLysGluGly.....LeuG 118
|||||:|||||:|||||:|||||:|||||:|||||:
594 GAATCACAGCTTTAATTACCGTTGGTAAAGACGCTAGCGTAAACCTTATTG 643
|||||:|||||:|||||:|||||:|||||:|||||:
118 lnGluGlyLeuLysAsnAlaGly..... 125
|||||:|||||:|||||:|||||:|||||:|||||:
644 GTGGCAAGGTGAAACGAGGGCGGTATGCGTAAATGGCGGTAGTATT 693
|||||:|||||:|||||:|||||:|||||:|||||:
126 ...LeuAlaValGlyGlnProLeuLysGlnAlaThrValGlnMetIleG 141
|||||:|||||:|||||:|||||:|||||:|||||:
694 TCTTTACTTGACGGGCAAAATTC.....ACCATCAGCGATATA... 732
|||||:|||||:|||||:|||||:|||||:|||||:
141 uThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyrTyrAsnThrGlu 158
|||||:|||||:|||||:|||||:|||||:|||||:
733 .....ATAAATCCCAACCA 745
|||||:|||||:|||||:|||||:|||||:|||||:
158 leThrValLysGlnThrMetLeuAspGlyAsnArgValLysLeuAspMet 174
|||||:|||||:|||||:|||||:|||||:|||||:

```

```

746 TCACTTACAGCATTTGCTGCACCTGAAAAACGAGCATCAATCTGGCGCAT 795
|||||:|||||:|||||:|||||:|||||:|||||:
175 ThrPheAlaGluGlyLysProAlaArgValValAspIleAsnIle..... 189
|||||:|||||:|||||:|||||:|||||:|||||:
796 ATTTTGGCCAAAGGTGGT.....AACATTAAATGTCGCCGC 830
|||||:|||||:|||||:|||||:|||||:|||||:
190 .....IleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValL 204
|||||:|||||:|||||:|||||:|||||:|||||:
831 TGCCACTATTTCGCAATAAAGGTAACCTTTCTGCGACTCTGTAAGC.... 876
|||||:|||||:|||||:|||||:|||||:|||||:
204 euAlaIleLysAspAsnLysIleAsn....ProLeuSerLysAlaAspArg 219
|||||:|||||:|||||:|||||:|||||:|||||:
877 .....AAAGATAAAGGTGAACATTCTCTCTGCCAAGAAGGT 918
|||||:|||||:|||||:|||||:|||||:|||||:
220 TyrThrGlnGlyLysLeuValThrSerLeuGluAsnLeuArgAlaLysTy 236
|||||:|||||:|||||:|||||:|||||:|||||:
919 GAAGCGGAATTTGCGGTGTAATTTCCGCTCAAAATCAGCAAGCC..... 963
|||||:|||||:|||||:|||||:|||||:|||||:
236 rLeuAsnAlaGlyPheValArgPheGluIleLysAspAlaLysLeuAsnI 253
|||||:|||||:|||||:|||||:|||||:|||||:
964 .....AAAGTGGTAAAGTTGATGA 982
|||||:|||||:|||||:|||||:|||||:|||||:
253 leAsnGluAspLysAsnArgIle.....PheValGluIleSer 265
|||||:|||||:|||||:|||||:|||||:|||||:
983 TTACAGCGCATAAAGTTTACATTGAAACGGGTGCAGTTATCGACCTTTTCG 1032
|||||:|||||:|||||:|||||:|||||:|||||:
266 LeuHisGlu...GlyGluGlnTyrArgPheGlyGlnThrGlnPheLeuG 281
|||||:|||||:|||||:|||||:|||||:|||||:
1033 GGTAAAGAGGGGGAGAACTTATCTTGGCGGTGACGAGCGTGGCGAAGG 1082
|||||:|||||:|||||:|||||:|||||:|||||:
281 yAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLeuLysPheLys 298
|||||:|||||:|||||:|||||:|||||:|||||:
1083 T.....AAAAACGGCATTCATTAAGCAAAAGAAACCACTT 1117
|||||:|||||:|||||:|||||:|||||:|||||:
298 laGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsnIle 314
|||||:|||||:|||||:|||||:|||||:|||||:
1118 TAGAAAAAGGC.....TCAACAATTAATGTG 1143
|||||:|||||:|||||:|||||:|||||:|||||:
315 SerThrLys.....PheGlyAsp..... 320
|||||:|||||:|||||:|||||:|||||:|||||:
1144 TCAGGTAAAGAAAAAGGTGGCGCGCTATTGTATGGGCGCATATTGGGTT 1193
|||||:|||||:|||||:|||||:|||||:|||||:
321 ....AspGlyTyrTyrAlaGlnIleArgProValThrArg..... 333
|||||:|||||:|||||:|||||:|||||:|||||:
1194 AATTGACGGCAATATTAAATGCCAAGGTAAGATATCGCTAAAACTGGTG 1243
|||||:|||||:|||||:|||||:|||||:|||||:
334 .....IleAsnAspGluSerArg 339
|||||:|||||:|||||:|||||:|||||:|||||:
1244 GTTTTGGAGACGTCGGGCGCATTTACTTATCCATTGATGATAAGCAATT 1293
|||||:|||||:|||||:|||||:|||||:|||||:
340 ThrValAspValGluTyrTyrIleAspPro.....ValHisPr 352
|||||:|||||:|||||:|||||:|||||:|||||:
1294 GTTAAACAAAGAAAGTGGCTACTAGACCCAGAGAAATGTGACTATTGAAGC 1343
|||||:|||||:|||||:|||||:|||||:|||||:
352 oValTyrValArgArgIleAsnPheThrGlyAsnPheLysThrGlnAsp. 368
|||||:|||||:|||||:|||||:|||||:|||||:
1344 TCCTTCGCTTCTCGGTGAGCTGGGTGCGGATAGGAATTCACACTCGG 1393
|||||:|||||:|||||:|||||:|||||:|||||:
369 ..GluValLeuArgArgGluMetArgGlnLeuGluGlyAlaLeuAlaSer 384
|||||:|||||:|||||:|||||:|||||:|||||:
1394 CAGAGGTGATAAAGATGACCCCTAAATAAATAACACCTCTCTTGACAA 1443
|||||:|||||:|||||:|||||:|||||:|||||:
385 AsnGlnLysIleGlnLeuSerArgAlaArgLeuMetArgThrGlyPhePh 401
|||||:|||||:|||||:|||||:|||||:|||||:
1444 CTAAACCAATACACCATTTCA.....AATCTTCTGAAAAGTGGCCACGT 1487
|||||:|||||:|||||:|||||:|||||:|||||:
401 eLysHisValThrValAspThrArgProValProAsnSerProAspGln 418
|||||:|||||:|||||:|||||:|||||:|||||:
1488 GGTGAACATAACGCAAGGAGAAACTTACCGTTAATAGCTCTATCATGA 1537
|||||:|||||:|||||:|||||:|||||:|||||:
418 aAsp.....ValAsnPheValValGluGlnProSerGlySerSer 432
|||||:|||||:|||||:|||||:|||||:|||||:
1538 TAGAAAAGAGGCTCCACCTTAATTTCTCCACAGTGAAGGTGAGGGCGGTCAA 1587
|||||:|||||:|||||:|||||:|||||:|||||:

```

```
alignment_scores:      Length: 882
                       Quality: 136.00
                       Ratio: 0.367
                       Gaps: 45
Percent Similarity: 42.063
Percent Identity: 18.141
```

alignment_block:

```

US-09-701-711-2 x US-09-074-658-69  ..
Align seg 1/1  to: US-09-074-658-69  from: 1  to: 2718

9  PheGlnValSerAlaMetThrMetAlaVal.....MetMetVa 21
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
31  TACCAAAACGACCCCTTACCTGCGCATCGCCAGTATTTTCTGCGGT 80
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
21  lMetSerThrHisAlaGlnAlaAlaPheMetAlaAsnAspIleThrI 38
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
81  GGTGATGACAGCGTGGCGTCTGATGACATCAGCGTCAATGCCACCAATG 130
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
38  leThrGlyLeuGlnArgValThrIleGluSerLeuGlnSerValLeuPro 54
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
131 TTACCCAACTGCCCAAGGACGGTTTCACCAATACCGAACACAGGTCTAT 180
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
55  PheArgLeuGlnValValSerGlu..... 63
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
181 GACAACACCAATACACCAACCAATCAGGGCAACAACACGGATACAGCAC 230
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
64  .....AsnGlnLeuAlaAspGlyValL 71
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
231 CAGCACAACTGACCCAAATAGCGGATACCACTGACACAAACACAA 280
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
71  ysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHisGln 87
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
281 AAATGCCCGCGCGAGGGTGTGATGGGTAAATTCGTGATACC 330
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
88  GluGlyArgIleIleTyrGlnValThrGluArgProIleAlaGluI 104
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
331 AGCGAAAAAATACCCAGATTTACC...AAAGATTACAGGACGCGT 377
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
104 eAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluGlyL 121
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
378 ACATACAGCAGG.....CAAGGCT 397
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
121 euLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThrVal 137
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
398 TACAGTACTTAGCCACCAAGACGCTGGCCAGATGGCAGGACAGGT 447
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
138 GlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTyr 154
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
448 AAAAAGCTTACGCCAGCCATCAGCAGTGATGACATTACACCACTTATTT 497
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
154 rAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgValL 171
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
498 TGAT..... 501
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
171 ysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAspIle 187
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
502 .....AAATCCCCAAAATATCGGATCTG 525
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
188 AsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspValLe 204
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
526 CACCTAGAAAAACGAGCATGTGTGTGATGCTAAAAAAGCAAAATACAT 575
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
204 u.....AlaIleLysAspAsnLysIleAsnProLeuS 215
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
576 CAAATATATATGTTATGGTGCATTTGTCATCACCTGCCAAAAGCCACCT 625
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
215 erLysAlaAspArgTyrThrGlnGluLysLeuValThrSerLeuGluAsn 231
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
626 ACATG.....AATTATCAACAAGAA.....CRAAAC 651
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
232 LeuArgAlaLys.....TyrLeuAsnAlaGlyPheValAr 243
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
652 ATCAAAAACAAAACACGCGGATGATTATCAAAACATTCGTTTGGCTA 701
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
243 gPheGluIleLysAspAlaLysLeuAsn.....IleAsnG 255
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
702 TATGGAGCTAAGAGCTGGACCTAAATAAAAAAAGGTGACAGACCCAGA 751
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

255 luAspLysAsnArg.....IlePheValGluIleSerLeu...HisGlu 268
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
752 GCGACAAGAACCGTGCCATCATTTTACCACACACCTATTTTATATCAT 801
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
269 GlyGluGlnTyr.....ArgPheGlyGlnThrGlnPheLe 280
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
802 GGTGAGAAATGCCAGCACCCATCTGCCAAAGCGGTAATTTTGACTATGA 851
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
280 uGlyAsnLeuThrTyrThrGlnAlaGluLeuGluAlaLeuLysPheL 297
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
852 GGGCAATTTGTTGTAT.....CTGACCGATGTCA 880
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
297 ysAlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAsnAsn 313
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
881 AA.....AAAGCCCATTTTAGATAAAACAGACAGATAAA 915
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
314 IleSerThrLysPhe.....GlyAsp..... 320
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
916 GTAGGCACTTATTTTAACTCAACCAGAAATCAAAATGAAGCGGATTTGGT 965
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
321 .....AspGlyTyrTyrTyrAlaGlnIleArgP 330
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
966 GAGTGCAGCACACATTTATCTAAACAGCTTTAAATATAAACACACACCCCGG 1015
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
330 roValThrArgIleAsnAspGluSerArgThrVal.....AspValGlu 344
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1016 CCAGTTATAGCGTGGACTTTGATCAAAATACCTTAAAGGCAAAATTTGCT 1065
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
345 TyrTyrIleAspProValHisProVal.....TyrValArgAr 357
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1066 TATTATGACAAACCAACAAAGCAACAGCCGATGGCGTTATATACAGA.. 1113
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
357 gIleAsnPheThrGlyAsnPheLysThrGlnAspGluValLeuArgArg 374
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1114 .....AGTCAGTTTGATACCGACAAAAGAGGTCAATGAAGCCG 1150
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
374 luMetArgGlnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeu 390
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1151 ATGTCTATGAGATTGACGCCCAAGATTAATGGCAACCGCTTTACTGGCACA 1200
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
391 SerArgAlaArgLeu.....MetArgThrGlyPhePheLysHisVa 404
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1201 GCCAAATCTTTGATGTATGATACACCAATACCGCCTTTTGTGTTAAAGA 1250
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
404 lThrValAspThrArgProValProAsnSerProAspGlnValAspVala 421
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1251 GCTGTTCTCCAAAAGCCCAATCCCAACACCCAGAC..... 1287
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
421 snPheValValGluGluGlnProSerGlySerSerThrIleAlaAlaGly 437
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1288 .....CCCAACTCAGATACGCTAGAGAGCGGG 1314
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
438 ...TyrSerGlnSerGlyGlyValThrPheGlnPheAspValSerGlnAs 453
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1315 TTTTATGTTGAGTCGGGC.....GATGAGCTGGCGGG 1346
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
453 nAsnPheMetGlyThrGlyLysHisValAsnAlaSerPhe..... 466
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1347 TAAATTTTATCCAATGAC.....AACGCAACTTTTGTGTGCTCTTTG 1387
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
467 .....SerArgSerGluThrArgGluValTyrSerLeuGlyMetThrAsn 481
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1388 GTGGCAACGAGACAAAACG.....ACCGAA 1413
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
482 ProTyrPheThrValAsnGlyValSerGlnSerLeuSerGlyTyrTyrAr 498
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1414 CTTGTCGCCACA..... 1425
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
498 gLysThrLysTyrAspAsnLysAsnIleSerAsnTyrValLeuAspSer 515
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1426 .AAAACGCTGTAT..... 1437
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```
515 yrGlyGlySerLeuSerTyrGlyTyrProIleAspGluAsnGlnArgIle 531
1438 .....TTT 1440
532 SerPheGlyLeuAsnAlaAspAsnThrLysLeuHisGlyGlyArgPheMe 548
1441 AGTACAGAGATTGAAACCCAGCAGCAGCTTTGTCGATGAAGAGAT 1490
548 tGlyIleSerAsnValLysGlnLeuMetAlaAspGlyGlyLysIle... 563
1491 TGGT .....AGCATTATTGACGGTAAAGATTAAATG 1522
564 .....GlnValAspAsnAsnGlyIleProAspPheLysHis 575
1523 ATGAAGTCAATAATCAATGAAGTGAACCTGTCCTGTCAGTAAATAA 1572
576 AspTyrThrThrTyrAsnAlaIleLeuGlyTyrPAsnTyrSerLeuAs 592
1573 GAATATTATGATAT .....AATTATGGACGA ..... 1599
592 pArgProValPheProThrGlnGlyMetSerHisSerValAspLeuThrV 609
1600 .....CCCAACAACAATACCAAAAAATAAAGCCAGCG 1636
609 aIGly .....PheGlyAspLysThrHisGlnLysValVal 620
1637 TCCAAAAAACCGTGTATTATTTGGTCAG .....CATGATAGTTTAT 1680
621 TyrGlnGlyAsnIleTyrArgProPheIleLys ..... 631
1681 TTTAATGGTAACATTATTATGACTTATCAGCCAAAGAAAGCAAAAGCTGG 1730
632 .....LysSerValLeuArgGlyTyr ..... 638
1731 TGTCCTCCCAAGATACAGCAGCACCATAAGAGTATTTGGCTAAATACCCAG 1780
639 ..AlaLysLeuGlyTyrGlyAsnAsnLeu ..... 647
1781 ATGCCAAAGTAAGCAGACAGACAATAAGTTACCAAAATCGTTCTACAACAA 1830
648 .....ProPheTyrGluAsnPheTyrAlaGlyGlyTyrGlySe 660
1831 GCCAAAGATAAGCG...TATACCGCATTCATGCC ..... 1863
660 rValArgGlyTyrAspGlnSerSerLeuGlyProArgSerGlnAlaTyrL 677
1864 .....AAAGCTATGACCACATCAGTTTGGT .....GAAGTATTGT 1900
677 euThrAlaArgArgGlyGlnGlnThrThrLeuGlyGluValValGlyGly 693
1901 ATAATGATAACAAAGGCAACCCCAACAGCAGATTATTTGTGCAAGCGGT 1950
694 AsnAlaLeuAlaThrPheGlySerGluLeuIleLeuProLeuPro ..... 708
1951 CAAGCG .....GATGTCAGTACTCAGCTGCCAGTCG 1982
709 .....PheLysGlyAspTrpIleAspGlnValArgProValI 721
1983 AGGTAATTCACCTATAATAGTCTTTGGCAGGCTAC ..... 2019
721 lePheIleGluGlyGlyGlnValPheAspThrThrGlyMetAspLysGln 737
2019 ..... 2019
738 ThrIleAspLeuThrGlnPheLysAspProGlnAlaThrAlaGluGlnAs 754
2020 .....CTGACCCAGAAAAAGACAAAGGTTTATAGCAAAAGATGAGGA 2060
754 nAla .....LysAlaAlaAsnArgProLeuLeuThrGlnAsp 766
2061 TACCATCAAGCAAAAGGCTCTTAAGATATATATTGACCAAGAC 2106
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq.us-08-841-178-25
```

```
seq_documentation_block:
; Sequence 25, Application US/08841178
; Patent No. 580275
; GENERAL INFORMATION:
; APPLICANT: Fischhoff, David A.
; TITLE OF INVENTION: Synthetic Plant Genes and Method for Preparation
; FILE REFERENCE: 38-21(15119)A
; CURRENT APPLICATION NUMBER: US/08/841.178
; CURRENT FILING DATE: 1997-04-29
; EARLIER APPLICATION NUMBER: US 08/433,111
; EARLIER FILING DATE: 1995-05-03
; EARLIER APPLICATION NUMBER: US 07/959,506
; EARLIER FILING DATE: 1992-10-09
; EARLIER APPLICATION NUMBER: US 07/476,661
; EARLIER FILING DATE: 1990-02-12
; EARLIER APPLICATION NUMBER: US 07/315,355
; EARLIER FILING DATE: 1989-02-24
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.0 - beta
; SEQ ID NO 25
; LENGTH: 3534
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:synthetic
; OTHER INFORMATION: structural gene encoding insecticidal protein
; OTHER INFORMATION: derived from B.t.k. HD-73
US-08-841-178-25

alignment_scores:
  Quality: 134.50      Length: 716
  Ratio: 0.414        Gaps: 35
  Percent Similarity: 45.391      Percent Identity: 19.832

alignment_block:
US-09-701-711-2 x US-08-841-178-25 ..
Align seg 1/1 to: US-08-841-178-25 from: 1 to: 3534

5 TyrPheLysGlyPheGlnValSerAla ..... 13
943 TACTGTGTCGACACAGATCATGGCTCTCCAGTTGGATTGAGTCAGCGGCC 992
14 ....MetThrMetAlaValMetMetValMetSerThrHisAlaGlnAlaA 29
993 CGAGTTTACCTTCTCTCTATGGAACCTATGGAAACGCCGCTCCACAAC 1042
29 laAspPheMetAlaAsnAspIleThrIleThrGlyLeuGlnArgValThr 45
1043 AAGCTATCGTTGCTCAA .....CTAGTTCAGGGTGTC 1074
46 IleGluSerLeuGlnSerValLeu .....ProPheArgLeuGlyG1 59
1075 TACAGAACCTTGTCTTCCACCTGTACAGAACGACCTTCAATATCGGT.. 1122
59 nValValSerGluAsnGlnLeuAla .....AspGlyValLysAlaLeuT 74
1123 ....ATCAACAACGACCAACTTCCGTTCTTGACGGAACAGAGTCGCCT 1168
74 yrAlaThrGlyAsnPheSerAspValGlnValTyrHisGlnGluGlyArg 90
1169 ATGGAACCTCTCTTAACCTTGCATCGCTGTTTACAGAAAGAGCGGAACC 1218
91 IleIleTyrGlnValThrGluArgProLeuIleAlaGluIleAsnPheG1 107
1219 GTT .....GATTCTTGGACGAAATCCCAACCACA 1247
107 uGlyAsnArgLeuIleProLysGlyGlyLeuGlnGluGlyLeuLysAsn 124
1248 GAACAACAATGTGCCACCCAGCAAGGATTCTCCACACAGGTTGAGCCACG 1297
```

```

2012 2012  AGTGCACACATCGCAAGCGACTCAGTGATGAACGC...AATTACTACTCAA 2058
398  ThrGlyPhePheLysHisValThrValAspThrArgProValProAsnSe 414
2059 2059  GATTCAAATTTCCANA... 2073
414  rProAspGlnValAspValAsnPheValValGluGluGlnProSerGlyS 431
2074 2074  ....GACATTAAT...AGCAACCCAGAA... 2094
431  erSerThrIleAlaLacGlyTyrSerGlnSerGlyGlyValThrPheGln 447
2095 2095  ....CGTGGGTGGGGCGGAAGTACAGGGATTACCATCCAA 2130
448  ....PheAspValSerGlnAsnAsnPheMetGlyThrGlyLysHisVa 462
2131 2131  GGAGGGGATGACGTATTAAAGAAAT... 2157
462  IasnAlaSerPheSerArgSerGluThrArgGluValTyrSerLeuGlyM 479
2157 2157  .... 2157
479  etThrAsnProTyrPheThrValAsnGlyValSerGlnSerLeuSerGly 495
2158 2158  ....TAGCTCACACATATCAGTACCTTTGTAGTGCTATCCA 2196
496  TyrTyrArgLysThrLysTyrAspAsn.....LysAsnIleSerAs 509
2197 2197  ACATATTGTATCAAAAAATCGATGAATCAAAATTAAGAAGCTTTTACCG 2246
509  nTyrValLeuAspSerTyrGlyGlySerLeuSerTyrGlyTyrProIleA 526
2247 2247  TTATCAATTAAAGGGGTAT.....ATCG 2269
526  spGluAsnGlnArgIleSerPheGlyLeuAsnAlaAspAsnThrLysLeu 542
2270 2270  AAGATGATGCAGAGACTTAGAAATCATTTAAATTCGCTACAATGCAGAAACAT 2319
543  HisGlyGlyArgPheMetGlyIleSerAsnValLysGlnLeuMetAlaAs 559
2320 2320  GAAACAGTAATATGCCAGGTACGGGTCTTATATGGCGCGTTTCAGGCCCA 2369
559  poly.....GlyLysIleGlnValAspAsnAsnGlyIleProAspPheL 574
2370 2370  AAGTCCAATCGGAAAGTGTGGAGAGCCGAATCGATCGCGCCACAC... 2415
574  yHisAspTyrThrThrTyrAsnAlaIleLeuGlyIleTyrAsnTyrSerSer 590
2416 2416  ....CTGTAATGGAAT...CTGTAC 2433
591  LeuAspArgProValPheProThrGlnGlyMetSerHis..... 603
2434 2434  TTAGATTCTTGTAGGGATGGAGAAAGTGTGCCCATCATTCGCATCA 2483
604  ....SerValAspLeuThrValGlyPheGlyAspLysThrHisGlnLysV 619
2484 2484  TTTCTCCTTAGACATGTATGTAGATGTACAGACTTAATAGAGGACCTAG 2533
619  alValTyrGlnGlyAsnIleTyrArgProPheIleLysSerValLeu 635
2534 2534  GTGTATGG.....GTGATCTTTAGATTAGACGCCAA 2565
636  ArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuProPheTyrGlu 651
2566 2566  GATGGCAGCAGCACTATAGG.....AATCTAGATTCTTCGAA 2604

```

; GENERAL INFORMATION:

3018 AATGATATATAAATACAGCCTTT.....AACCTAAAAATAATAA 3058
50 nSerValLeuProPheArgLeuGlyGlnValValSerGluAsnGlnLeuA 67
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3059 CAACCCATTGACTTT.....GTCTCCACTTATGACATTG 3093
67 laaspGlyValLysAlaLeuTyrAlaThrGlyAsnPheSerAspValGln 83
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3094 TTGAC.....TTTGCCCAATGCCAATGCCACCCGCGACA 3128
84 ValTyrHis.....GlnGluGlyArgIleTyrGlnValTh 96
||| ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3129 GTAAACCCATGATACCGCTACAAAACACAGTAAGAGTGATATATGTGAA 3178
96 rGluArgProLeuIleAlaGluIleAsnPheGluGlyAsnArgLeuIleP 113
:
3179 T..... 3179
113 rLysGluGlyLeuGlnGluGlyLeuLysAsnAlaGlyLeuAlaValGly 129
3179 3179
130 GlnProLeuLysGlnAlaThrValGlnMetIleGluThrGluLeuThrAs 146
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3180GTGGATGATACACCACTCATCTAACAGGCACGTAT..... 3215
146 nGlnTyrIleSerGlnGlyTyrTyrAsnThrGluIleThrValLysGlnT 163
||| ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3216GACAAATAAAACCTTGGCGGTCAAAACCA 3243
163 hrMetLeuAspGlyAsnArgValLysLeuAspMetThrPheAlaGluGly 179
|| ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3244 CC.....AAACTGAACAAAACAAAGTGCTTAATGGT 3272
180 LysProAlaArgValValAspIleAsnIleGlyAsnGlnHisPheSe 196
::: ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3273 AATACAGCACTAACTTTAATGTTAAC.....TCTAG 3304
196 rAspAlaAsp.....LeuIleAspValLeuAlaIleLysAspAsnLysIleA 212
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3305 TGATGAAGATGCCCTTGTAAAGCCCAACACATCCCGCAAAAT.....CTAA 3351
212 snProLeuSerLysAlaAspArgTyrThrGlnGluLysLeuValThrSer 228
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3352 ACACCTAGCCAAAGGAATTCACACCACCAAGGACGACACACCGCC 3401
229 LeuGluAsnLeuArgAlaLysTyrLeu.....As 238
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3402 CTCAAAACCTTTACCGTTTAAAGGTAGATGAAATAATAATGCTGTGATGA 3451
238 nAlaGlyPheValArgPheGluIleLysAspAlaLys..... 250
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3452 CGCCACGCCATCACCGTGGTCAAAAGAACGCAATAATCAAGTCAACA 3501
251LeuAsnIleAsnGluAspLysAsn 258
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3502 CCCTAACACCTCAAAAGGTGAAACCGGTCTTAATAATAAACCCACCAAAAT 3551
259 ArgIlePheValGluIleSerLeuHisGluGlyGluGlnTyrArgPheG 275
3551 3551
275 yGlnThrGlnPheLeuGlyAsnLeuThrTyr.....ThrGlnAlaG 289
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3552GGTACGGTTACTTGGCATTAACACCAACGCG 3585
289 luLeuGluAlaLeuLeuLysPheLysAlaGluGluGlyPheSerGlnAla 305
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3586 GTCTTAAAGCGGCAAAAGCACCCCTAAACACCGGTGGCTGTCT..... 3629
306 MetLeuGluGlnThrThrAsnAsnIleSerThrLysPheGlyAspAspG 322
||| :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
3630ATTAAGCCCTGCTAGCGAACAATCCAAAGTCGCTGCTGATGG 3676

[illegible]

4805 AGAAGTTGCCAAAGACAAACTGGTCCCAAGCCCAA..... 4841
 772 erAlaglyValGlyAlaThrTPTyThrProileGlyProLeuSer... 787
 4842ACCCAGATGGCACATTGGCTCAA 4865
 788 IleSerTyAlaLysProLeuAsnLysLysGln...AsnAsp 800
 4866 ATGAATGTCAATCAGTCATTAAACAAGAACAAAGTAATGAT 4907

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-671-817A-1

seq_documentation_block:
 ; Sequence 1, Application US/07671817A
 ; Patent No. 5424409
 ; GENERAL INFORMATION:
 ; APPLICANT: Ely, Susan
 ; APPLICANT: Tippet, Janet M
 ; TITLE OF INVENTION: DNA constructs
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Cushman, Darby and Cushman
 ; STREET: Eleventh floor, 1615 L Street, N.W.
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20036-3601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/671,817A
 ; FILING DATE: 19910401
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: GB 8823068.5
 ; FILING DATE: 30-SEP-1988
 ; PRIOR APPLICATION DATA: PCT/GB89/01157
 ; FILING DATE: 29-SEP-1989
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202) 861-3000
 ; TELEFAX: (202) 822-0944
 ; TELEX: 6714627 CUSH
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2990 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-07-671-817A-1

alignment_scores:
 Quality: 130.00 Length: 720
 Ratio: 0.406 Gaps: 35
 Percent Similarity: 44.444 Percent Identity: 19.583

alignment_block:

US-09-701-711-2 x US-07-671-817A-1 ..
 Align seg 1/1 to: US-07-671-817A-1 from: 1 to: 2990
 5 TyrPheLysGlyPheGlnValSerAlaMetThrMetAla..... 17
 1023 TATTGGTCAGGCGCATCAATATGCTTCTCTCGGGTTTCGGGGCC 1072
 18ValMetMetValMetSerThrHisAlaGlnAlaAspPheM 32
 1073 AGAATTCAGTTTCGGCTATATGGAACCATGGAAATGCAGTCCA.... 1118

32 etAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleGluSer 48
 1119CAACAAGTATTGTGTCTCAACTA 1142
 49 LeuGlnSerValLeu.....Proph 55
 1143 GGTGAGGCGGTGTATAGAACATTATCTCTACTCTTTTATAGAAGACCTTT 1192
 55 eArgLeuGlyGlnValValSerGluAsnGlnLeuAla.....AspGlyV 70
 1193 TAATATAGG.....ATAAATAACCAACTATCTGTCTTGACGGGA 1236
 70 alLysAlaLeuTyAlaThrGlyAsnPheSerAspValGlnValTyrHis 86
 1237 CAGATTTGCTTATGGAACCTCTCAAAATTTGCCATCCGCTGTATACAGA 1286
 87 GlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaG 103
 1287 AAAAGCGGAACGGTA.....GATTCCGCTGGATGA 1315
 103 uLeAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
 1316 AATACCACACAGATAACACGTCGCCCTAGGCAAGGATTTAGTCATC 1365
 120 lLysLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThr 136
 1366 GATTAAAGCCAT.....GTTTCAATGTTTCGTTACGGCTCTAGTAGTAGT 1409
 137 ValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTy 153
 1410 GTAAGTATAATAAGAGCTCCTCTCTCTGGATACATCGTAGTGTCTGA 1459
 153 rTyAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgV 170
 1460 ATTTAAT...AATATAATTGTCATCGGATAGTATTACTCAAAATCCCTGCAG 1506
 170 alLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAsp 186
 1507 TGAAGGGAACCTTCTTTT.....AAT 1529
 187 IleAsnIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVa 203
 1530 GGTTCGTGTAATTCAGGACCAGGATTTACTGGTGGGACTTAGTTAGATT 1579
 203 lLeuAlaIleLysAspAsnLysIleasn.....proL 214
 1580 AAATAGTAGTGGAAATAACATTCAGAATAAGAGGTATATTGAAGTTCCAA 1629
 214 eu.....SerLysAlaAspArgTyThrGlnGluLysLeuValThr 227
 1630 TTCACCTCCCATCGACATCTACCAGATAT..... 1658
 228 SerLeuGluAsnLeuArgAlaLysTyThrLeuAsnAlaGlyPheValArgPh 244
 1659CGAGTTCTGTACGGTATGCTTCT..... 1682
 244 eGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIleP 261
 1683GTAACCCCGATTCACCTCAACGTTAAATTTGGGGTAATTCATCCATTT 1728
 261 he.....ValGluIleSerLeuHisGluGlyGln 271
 1729 TTTCCAATACAGTACAGCTACAGCTACGTCAATAGATANTCTACANTCA 1778
 272 TyrArgPheGlyGln.....ThrGlnPheLeuG 281
 1779 AGTGATTTTGGTTATTTCGAAAGTGCCAATGCTTTTACATCTTCATTAGS 1828
 281 yAsnLeuThrTyrThrGln..... 287
 1829 TAATATAGTAGTGTAGAAATTTTAGTGGGACTGCAGGAGTGAATAAG 1878

```
288 .....AlaGluLeuGluAlaLeuLeuLys 295
1879 ACAGATTGATTTATTCACACTCGAGCTGATATAAT 1928
296 PheLysAlaGluGluGlyPheSerGlnAlaMetLeuGluGlnThrAs 312
1929 CTGGAAGAGCGCAGAGCGGTGAATCGCTGTTT...ACGCTACAAA 1975
312 nasIleSerThrLysPheGlyAspAspGlyTyrTyrAlaGlnIleA 329
1976 CCAACTAGGCGTAAAA..... 1991
329 rgProValThrArgIleAsnAspGluSerArgThrValAspValGluTyr 345
1992 .....ACAAATGTACGGAT.....TAT 2009
346 TyrIleAspProValHisProValTyrValArgArgIleAsnPheThrG 362
2010 CATATTGATCAAGTG.....TC 2026
362 yAsnPhelysthr.....GlnAspGluValLeuArgArgGluMetArgG 377
2027 CAATTAGTTACGTATTATTCGGATGAATTTGTCTGGATGAAAGCGAG 2076
377 InLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAla 393
2077 AATTGTCGAGAAGTCAACATCGGAGCGACTCAGTGATGAACGC... 2123
394 ArgLeuMetArgThrGlyPhePheLysHisValThrValAspThrArgPr 410
2124 AATTACTCCAAGATTCAAAATTTCAAA..... 2150
410 ovalProAsnSerProAspGlnValAspValAsnPheValValGluGluG 427
2151 .....GACATTAAT.....AGGC 2163
427 InProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSerGlyGly 443
2164 AACAGAAA.....CGTGGTGGCGGAGGACACAGG 2195
444 ValThrPheGln.....PheAspValSerGlnAsnAsnPheMetGlyTh 458
2196 ATTACCATCCAAGAGGGATGAGCTATTAAAGAAAAT..... 2234
458 rGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluValT 475
2234 ..... 2234
475 yrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerGln 491
2235 .....TACGTACACATATCAGGTACCTTGAT 2261
492 SerLeuSerGlyTyrTyrArgLysThrLysTyrAspAsn.....Ly 505
2262 GAGTGTATCCACATATTTGTATCAAAAATCGATGAATCAAAATTA 2311
505 sasIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrG 522
2312 AGCCTTTACCGTTATCAATTAACAGGGTAT..... 2342
522 lyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 538
2343 .....ATCGAAGATAGTCAAGACTAGAAATCTATTTAATTCGCTAC 2384
539 AsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysG 555
2385 ANTGCARAAACATGAACAGTAAATGTGCCAGGTACGGGTTCCTTATGCC 2434
555 leuMetAlaAspGly.....GlyLysIleGlnValAspAsnAsnGlyI 570
2435 GCTTTGACCCCAAGTCCAATCGAAAGTGTGGAGAGCGGAATCGATCG 2484
570 leProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTyr 586
```

```
2485 CGCCACAC.....CTTGAATGG 2501
587 AsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerH 603
2502 AAT...CCTGACTTAGATTGCTGCTAGGATGGAGAAAAGTGTGCCCA 2548
603 s.....SerValAspLeuThrValGlyPheGlyAspLysT 615
2549 TCATTCGCATCATTTCTCCTTAGACATGATGATAGGATGTACACACTAA 2598
615 hrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLys 631
2599 ATGAGGACCTAGGTGTATGG.....GTGATCTTTAAG 2630
632 LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuPr 648
2631 ATTAAGACGCAAGATGGCGACGCAAGACTAGGG.....AATCTAGA 2671
648 oPheTyrGlu 651
2672 GTTCTCGAA 2681
seq_name: /cqn2_6/ptodata/2/lna/5A_COMB.seq:US-07-828-788A-7
seq documentation block:
; Sequence 7, Application US/07828788A
; Patent No. 5273746
; GENERAL INFORMATION:
; APPLICANT: PAYNE, JEWEL M.
; APPLICANT: HICKLE, LESLIE A.
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES
; TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID R. SALIWANCHIK
; STREET: 2421 N.W. 41ST STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/828,788A
; FILING DATE: 19920129
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SALIWANCHIK, DAVID R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MA75
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-372-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3531 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; STRAIN: KURSTAKI
; INDIVIDUAL ISOLATE: PS81GG
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK
; CLONE: 81GG
```



```

:::      |||      |||      |||      |||      |||
2182 GAGTGCTATCAACATATTTGTATCAAAAATCGATGAATCAAAATATAA 2231
505 sAsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrG 522
2232 AGCCTTACCGGTATCAATTAAGAGGTAT..... 2262
522 lyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 538
2263 .....ATCGAAGATAGTCAAGACTTAGAAATCTATTTAAATTCGCTAC 2304
539 AsnThrIysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysG 555
2305 AATGCAAAACATGAACAGTAATATGTCCAGGTACGGGTCCCTTATGGCC 2354
555 nLeuMetAlaSpGly.....GlyLysIleGlnValAspAsnAsnGlyI 570
2355 GCTTTCAGCCCAAGTCAATCGGAAAGTGTGGAGAGCGGAATCGATGCG 2404
570 leProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyT 586
2405 CGCCACAC.....CTTGAATGG 2421
587 AsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHi 603
2422 AAT...CCTGACTTAGATTGTCGTAGGATGGAGAAAGTGTGCCCA 2468
603 s.....SerValAspLeuThrValGlyPheGlyAspLysT 615
2469 TCATTCGCATCATTTCTCTTAGACATGTAGGTAGATGTACAGACTTAA 2518
615 hrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLys 631
2519 ATGAGGACTAGGTATGG.....GTGATCTTTAAG 2550
632 LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuPr 648
2551 ATTAAGACGCAAGATGGCAGCAAGACTAGG.....AATCTAGA 2591
648 oPheTyrGlu 651
2592 GTTCTCGAA 2601

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-920-085-1
seq_documentation_block:
; Sequence 1, Application US/07920085
; Patent No. 535661
; GENERAL INFORMATION:
; APPLICANT: PAYNE, JEWEL M.
; APPLICANT: SICK, AUGUST J.
; APPLICANT: THOMPSON, MARK
; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATE DENOTED B.T.
; TITLE OF INVENTION: PS81GG, ACTIVE AGAINST LEPIDOPTERAN PESTS, AND A GENE
; TITLE OF INVENTION: ENCODING A LEPIDOPTERAN-ACTIVE TOXIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROMAN SALIWANCHIK
; STREET: 2421 N.W. 41st STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/920,085
; FILING DATE: 19920727
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
```

```

1330 GTAAGTATTAATAGAGCTCCTATGTCTCTTGGATACATCGTAGTCTGA 1379
153 rTyrAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgV 170
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1380 ATTTAAT...AATATAATTGCATCGGATAGTATTACTCAATCCCTGCAG 1426
170 aLysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValAsp 186
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1427 TGAAGGGAACACTTCTTTT...AAT 1449
187 IleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVa 203
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1450 GGTTCGTGTAATTCAGACAGGATTTACTGTGGGACTTAGTAGATT 1499
203 lLeuAlaIleLysAspAsnLysIleAsn...ProL 214
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1500 AAATAGTAGTGGAAATAACATTTCAGATAGAGGGTATATTGAAGTCCAA 1549
214 eu.....SerLysAlaAspArgTyrThrGlnGluLysLeuValThr 227
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1550 TTCACCTTCCATCGACATCTACCAGATAT... 1578
228 SerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPh 244
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1579 .....CGAGTTCGTGTACGGTATGCTTCT... 1602
244 eGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIlep 261
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1603 ....GTAACCCCGATTCACCTCACGTTAATTGGGGTAATTCATCCATTT 1648
261 he.....ValGluIleSerLeuHisGluGlyGluGln 271
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1649 TTTCCAATACAGTACCAGCTACAGCTACGTCATAGATAATACATAACA 1698
272 TyrArgPheGlyGln.....ThrGlnPheLeuGlu 281
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1699 AGTGATTTTGGTATTGTGAAAGTGCCAAATGCTTTTACATCTTCATAGG 1748
281 yAsnLeuThrTyrThrGln..... 287
1749 TAATATAGTAGGTGTAGAAATTTAGTGGGACTCGAGGCTGAATAATAG 1798
288 .....AlaGluLeuGluAlaLeuLeuLys 295
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1799 ACAGATTTGAATTTATTCCAGTTACTGCAACACATCGAGGCTGAATAAT 1848
296 PheLysAlaGluGlyPheSerGlnAlaMetLeuGluGlnThrThrAs 312
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1849 CTGGAAGAGCGCAGAAGCGGTGAATGCGCTGTTT...ACGTCTACAAA 1895
312 nAsnIleSerThrLysPheGlyAspGlyTyrTyrAlaGlnIleA 329
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1896 CCAACTAGGGCTAANA..... 1911
329 rgProValThrArgIleAsnAspGluSerArgThrValAspValGluTyr 345
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1912 .....ACAAATGTACGGAT.....TAT 1929
346 TyrIleAspProValHisProValTyrValArgAlaIleAsnPheThrGl 362
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1930 CATATTGTCAAGTG.....TC 1946
362 yAsnPheLysThr.....GlnAspGluValLeuArgArgGluMetArgG 377
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1947 CAATTAGTTACGTATTATTCGGATGAATTTGTCTGGATGAAGCGAG 1996
377 lLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAla 393
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1997 AATTGTCCGAGAAAGTCAAAACATGCGAAGCGACTCAGTGATGAAGCC... 2043
394 ArgLeuMetArgThrGlyPhePheLysHisValThrValAspThrArgPr 410
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2044 AATTACTCCAAGATTCAAATTTCAA..... 2070

```

```

410 oValProAsnSerProAspGlnValAspValAsnPheValValGluGluG 427
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2071 .....GACATTAAAT.....AGGC 2083
427 lNProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSerGlyGly 443
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2084 AACCAAGAA.....CGTGGTGGGCGGAAGTACAGG 2115
444 ValThrPheGln.....PheAspValSerGlnAsnAsnPheMetGlyTh 458
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2116 ATTACCATCCAGAGGGGATGACGTATTTAAAGAAAT..... 2154
458 rGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluVal 475
2154 ..... 2154
475 yrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerGln 491
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2155 .....TAGCTCACACTATCAGTAGCTTTGAT 2181
492 SerLeuSerGlyTyrTyrArgLysThrLysTyrAspAsn.....Ly 505
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2182 GAGTGTATCCCAACATATTGTATCAAAAAATCGATGAATCAAAATPAAA 2231
505 sAsnIleSerAsnTyrValLeuAspSerTyrGlyGlySerLeuSerTyrG 522
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2232 AGCCTTTACCGTTATCAATTAAGAGGGTAT..... 2262
522 lyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 538
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2263 .....ATCGAAGATAGTCAAGACTTAGAAATCTATTAAATTCGCTAC 2304
539 AsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGl 555
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2305 AATGCAAAACATGAACAGATAATGTGCCAGGTACGGGTTCTTATGGCC 2354
555 nLeuMetAlaAspGly.....GlyLysIleGlnValAspAsnAsnGlyI 570
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2355 GCCTTCAGCCCAAGTCCAATCGAAAGTGTGGAGAGCCGAATCGATCGC 2404
570 leProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyTrp 586
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2405 CGCCACAC.....CTTGAATGG 2421
587 AsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHi 603
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2422 AAT...CCTGACTTAGATTGTCGTAGGGATGGAGAAAGTGTGCCCA 2468
603 s.....SerValAspLeuThrValGlyPheGlyAspLysT 615
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2469 TCATTCGCATCATTTCTCTTAGACATGTAGGTAGGTACAGACTTAA 2518
615 hrHisGlnLysValValTyrGlnGlyAsnIleTyrArgPropheIleLys 631
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2519 ATGAGGACCTTAGGTGTATGG.....GTGATCTTTAAG 2550
632 LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuPr 648
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2551 ATTAAGCCGCAAGATGGCCAGCAGACTAGGG.....AATCTAGA 2591
648 oPheTyrGlu 651
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2592 GTTCTCGAA 2601

```

seq_name: /cgn2_6/ptodata/2/lna/PTUS_COMB.seq: PCT-US92-11337-7

seq_documentation_block:

; Sequence 7, Application PC/TUS9211337

; GENERAL INFORMATION:

; APPLICANT: PAYNE, JEWEL M.

; APPLICANT: HICKLE, LESLIE A.

; TITLE OF INVENTION: NOVEL BACILLUS THURINGIENSIS ISOLATES

```
; TITLE OF INVENTION: ACTIVE AGAINST PHTHIRAPTERA PESTS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DAVID R. SALIWANCHIK
; STREET: 2421 N.W. 41st STREET, SUITE A-1
; CITY: GAINESVILLE
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/11337
; FILING DATE: 19921231
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 97/828,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: SALIWANCHIK, DAVID R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: MAY5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3531 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: BACILLUS THURINGIENSIS
; STRAIN: KURSTAKI
; INDIVIDUAL ISOLATE: PS81GG
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDA GEM (TM) - 11 LIBRARY OF AUGUST SICK
; CLONE: 81GG
; PCT-US92-11337-7

alignment_scores:
  Quality: 130.00      Length: 720
  Ratio: 0.406         Gaps: 35
  Percent Similarity: 44.444 Percent Identity: 19.583

alignment_block:
US-09-701-711-2 x PCT-US92-11337-7 ..

Align seg 1/1 to: PCT-US92-11337-7 from: 1 to: 3531

5 TyrPheLysGlyPheGlnValSerAlaMetThrMetAla.....17
|||||
943 TATTGGTCAGGCATCAATATATGCTTCTCCTGTCGGTTTTCGGGGCC 992
|||||
18 .....ValMetMetValMetSerThrHisAlaGlnAlaAlaAspPheM 32
|||||
993 AGAATTCACGTTCCGCTATATGGAACCATGGGAATGCAGCTCCA....1038
|||||
32 etAlaAsnAspIleThrIleThrGlyLeuGlnArgValThrIleGluSer 48
|||||
1039 .....CAACACGATTTGTTGCTCAACTA 1062
|||||
49 LeuGlnSerValLeu.....ProPh 55
|||||
1063 GGTACGGCGGTGTATAGAACATTATCTCTACTTTTATAGAACCTTT 1112
|||||
```

```
55 eArgLeuGlyGlnValValSerGluAsnGlnLeuAla.....AspGlyV 70
|||||
1113 TAATATAGG.....ATAAATAATCAACACTATCTGTTCTTACGGGA 1156
|||||
70 allysAlaLeuTyrAlaThrGlyAsnPheSerAspValGlnValTyrHis 86
|||||
1157 CAGAAATTTGCTTATGGAACTCTCAAAATTTGCCATCCGCTGTATACAGA 1206
|||||
87 GlnGluGlyArgIleIleTyrGlnValThrGluArgProLeuIleAlaG1 103
|||||
1207 AAAAGCGGAACGGTA.....GATTGCTGGATGA 1235
|||||
103 uileAsnPheGluGlyAsnArgLeuIleProLysGluGlyLeuGlnGluG 120
|||||
1236 AATACCACACAGAATAACAACGTCGCCACCTAGGCAAGGATTTAGTCATC 1285
|||||
120 lYleuLysAsnAlaGlyLeuAlaValGlyGlnProLeuLysGlnAlaThr 136
|||||
1286 GATTAAAGCCAT.....GTTTCAATGTTTCGTTACAGGCTCTAGTAGTAGT 1329
|||||
137 ValGlnMetIleGluThrGluLeuThrAsnGlnTyrIleSerGlnGlyTy 153
|||||
1330 GTAAGTATAATAAGAGCTCTATGTTCTCTGGATACATCTAGTAGTCTGA 1379
|||||
153 rTyAsnThrGluIleThrValLysGlnThrMetLeuAspGlyAsnArgV 170
|||||
1380 ATTTAAT...AATAAATGTCATCGGATAGTATTACTCAATCCCTGCAG 1426
|||||
170 allysLeuAspMetThrPheAlaGluGlyLysProAlaArgValValasp 186
|||||
1427 TGAAGGGAACCTTCTCTTT.....AAT 1449
|||||
187 IleAsnIleIleGlyAsnGlnHisPheSerAspAlaAspLeuIleAspVa 203
|||||
1450 GGTCTCTAATTTACAGACGAGGATTTACTGTGGGACTTAGTTAGATT 1499
|||||
203 IleuAlaIleLysAspAsnLysIleAsn.....ProL 214
|||||
1500 AAATAGTAGTGAAATATACATTCAGAAATAGAGGGTATATTGAAGTTCCA 1549
|||||
214 eu.....SerLysAlaAspArgTyrThrGlnGluLysLeuValThr 227
|||||
1550 TTCACCTTCCCATCGACATCTACAGATAT..... 1578
|||||
228 SerLeuGluAsnLeuArgAlaLysTyrLeuAsnAlaGlyPheValArgPh 244
|||||
1579 .....CGAGTTGCTGTCAGGTATGCTTCT..... 1602
|||||
244 eGluIleLysAspAlaLysLeuAsnIleAsnGluAspLysAsnArgIleP 261
|||||
1603 ....GTAACCCCGATTCACCTCAACGTTAATTGGGGCTAATTCATCATTT 1648
|||||
261 he.....ValGluIleSerLeuHisGluGlyGluGln 271
|||||
1649 TTTTCCAATACAGTACCAGCTACAGCTACGCTACGCTATAGATAATCTACAATCA 1698
|||||
272 TyrArgPheGlyGln.....ThrGlnPheLeuG1 281
|||||
1699 AGTGATTTTGGTTATTTTGAAGTGCCAATGCTTTTACATCTCATTAGG 1748
|||||
281 yAsnLeuThrTyrThrGln..... 287
|||||
1749 TAATATAGTAGGTGTAGAAATTTTAGTGGACTGCAGGAGTGATAATAG 1798
|||||
288 .....AlaGluLeuGluAlaLeuLeuLys 295
|||||
1799 ACAGATTTGAATTTATTCAGTTACTGCAACACTCGAGGCTGAATAAT 1848
|||||
296 PheLysAlaGluGlyPheSerGlnAlaMetLeuGlnThrThrAs 312
|||||
1849 CTGGAAGACGCGAGAGGGGTGAATCGCTGTTT...ACGCTACAAA 1895
|||||
312 nAsnIleSerThrLysPheGlyAspAspGlyTyrTyrTyrAlaGlnIleA 329
|||||
```



```

1896 CCAACTAGGCTAAA..... 1911
329 rgProValThrArgIleAsnAspGluSerArgThrValaspValGluTyr 345
1912 .....ACAAATGTAACGGAT.....TAT 1929
346 TyrIleAspProValHisProValTyrValArgArgIleAsnPheThrGI 362
1930 CATATTGATCAAGTG.....TC 1946
362 yAsnPheLysThr.....GlnAspGluValLeuArgGluMetArg 377
1947 CAATTTAGTTACGTATTATTTCGGATCAATTTCTCTGGATGAAAAGCGAG 1996
377 lnLeuGluGlyAlaLeuAlaSerAsnGlnLysIleGlnLeuSerArgAla 393
1997 AATTGTCCGAGAAAGTCAACATGCGAAGCGACTCAGTGATGAACGC... 2043
394 ArgLeuMetArgThrGlyPhePheLysHisValThrValAspThrArgPr 410
2044 AATTACTCAAGATTCAAATTTCAAA..... 2070
410 oValProAsnSerProAspGlnValaspValAsnPheValValGluGlu 427
2071 .....GACATTAAT.....AGGC 2083
427 lnProSerGlySerSerThrIleAlaAlaGlyTyrSerGlnSerGlyGly 443
2084 ACCAGAAA.....CGTGGGTGGGGCGGAGTACAGGG 2115
444 ValThrPheGln.....PheAspValSerGlnAsnAsnPheMetGlyTh 458
2116 ATTACCATCAAGAGGGGATGACGTATTTAAGAAAAAT..... 2154
458 rGlyLysHisValAsnAlaSerPheSerArgSerGluThrArgGluValt 475
2154 ..... 2154
475 yrSerLeuGlyMetThrAsnProTyrPheThrValAsnGlyValSerGln 491
2155 .....TAGCTCACACTATCAGGTACCTTTGAT 2181
492 SerLeuSerGlyTyrTyrArgLysThrLysTyrAspAsn.....Ly 505
2182 GAGTGTCTCCAAACATATTGTATCAAAAAATCGATGAATCAAAATTA 2231
505 sAsnIleSerAsnTyrValLeuAspSerTyrGlySerLeuSerTyrG 522
2232 AGCCTTTACCCGTTATCAATTAAGAGGGTAT..... 2262
522 lyTyrProIleAspGluAsnGlnArgIleSerPheGlyLeuAsnAlaAsp 538
2263 .....ATCGAAGATAGTCAAGACTTAGAAATCTATTAAATTCGCTAC 2304
539 AsnThrLysLeuHisGlyGlyArgPheMetGlyIleSerAsnValLysGI 555
2305 AATGCAAAACATGAAACAGTAAATGTCAGGTAACGGTTCCTTATGGCC 2354
555 nLeuMetAlaAspGly.....GlyLysIleGlnValAspAsnAsnGlyI 570
2355 GCTTTCAGCCCAAGCCCAATCGGAAAGTGTGGAGAGCCCAATCGATGCG 2404
570 leProAspPheLysHisAspTyrThrThrTyrAsnAlaIleLeuGlyT 586
2405 CGCCACAC.....CTTGAATGG 2421
587 AsnTyrSerSerLeuAspArgProValPheProThrGlnGlyMetSerHI 603
2422 AAT...CCTGACTTAGATTGTCGTAGGGATGGAGAAAAAGTGTGCCCA 2468
603 s.....SerValAspLeuThrValGlyPheGlyAspLysT 615

```

```

2469 TCATTGCGCATCATTTCTCCTTAGACATTGATGTAGGATGTACAGACTTAA 2518
615 hrHisGlnLysValValTyrGlnGlyAsnIleTyrArgProPheIleLys 631
2519 ATGAGGACCTAGGTGTATGG.....GTGATCTTTAAG 2550
632 LysSerValLeuArgGlyTyrAlaLysLeuGlyTyrGlyAsnAsnLeuPr 648
2551 ATTAAGACGCAAGATGGGCACGCAAGACTAGGG.....AATCTAGA 2591
648 opheTyrGlu 651
2592 GTTCTCGAA 2601

```

